NASA CONTRACTOR REPORT

NASA CR-11

NASA CR-1126

PRACTICAL RELIABILITY

Volume I - Parameter Variations Analysis

Prepared by
RESEARCH TRIANGLE INSTITUTE
Research Triangle Park, N. C.
for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JULY 1968

PRACTICAL RELIABILITY

Volume I - Parameter Variations Analysis.

Prepared under Contract No. NASw-1448 by
RESEARCH TRIANGLE INSTITUTE
Research Triangle Park, N.C.

for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

FOREWORD

The typical few-of-a-kind nature of NASA systems has made reliability a premium even on the initial items delivered in a program. Reliability defined and treated on the basis of percentage of items operating successfully has much less meaning than when larger sample sizes are available as in military and commerical products. Reliability thus becomes based more on engineering confidence that the item will work as intended. The key to reliability is thus good engineering--designing reliability into the system and engineering to prevent degradation of the designed-in reliability from fabrication, testing and operation.

This PRACTICAL RELIABILITY series of reports is addressed to the typical engineer to aid his comprehension of practical problems in engineering for reliability. In these reports the intent is to present fundamental concepts on a particular subject in an interesting, mainly narrative form and make the reader aware of practical problems in applying them. There is little emphasis on describing procedures and how to implement them. Thus there is liberal use of references for both background theory and cookbook procedures. The present coverage is limited to five subject areas:

Vol. I. - Parameter Variation Analysis describes the techniques for treating the effect of system parameters on performance, reliability, and other figures-of-merit.

Vol. II. - Computation considers the digital computer and where and how it can be used to aid various reliability tasks.

Vol. III. - Testing describes the basic approaches to testing and emphasizes the practical considerations and the applications to reliability.

Vol. IV. - Prediction presents mathematical methods and analysis approaches for reliability prediction and includes some methods not generally covered in texts and handbooks.

Vol. V. - Parts reviews the processes and procedures required to obtain and apply parts which will perform their functions adequately.

These reports were prepared by the Research Triangle Institute, Research Triangle Park, North Carolina 27709 under NASA Contract NASw-1448. The contract was administered under the technical direction of the Office of Reliability and Quality Assurance, NASA Headquarters, Washington, D. C. 20546 with Dr. John E. Condon, Director, as technical contract monitor. The contract effort was performed jointly by personnel from both the Statistics Research and the Engineering and Environmental Sciences Divisions. Dr. R. M. Burger was technical director with W. S. Thompson serving as project leader.

This volume is Vol. I. - Parameter Variation Analysis of the series. There has been a notable increase in interest on this subject due to higher precision requirements for system operation and optimization of system figures-of-merit such as performance, reliability, and effectiveness. Dr. R. A. Evans is the principal author of this report.

ABSTRACT

Concepts such as model, randomness, statistical ignorance, and statistical independence are explored and explained. The effects of variations of parameters can be evaluated by changing components in a physical model or by creating a conceptual/mathematical model of the system and then analyzing it. There are only a very few basic techniques for analysis of mathematical models and these are rather extensively treated. The uses to which these models and their analyses may be put are many and only a few of these are treated such as sensitivity and worst-case analyses, and calculation where extreme extrapolation is necessary. There is a brief discussion of the sources and uses of variations data both in purchased and in manufactured items. A series of appendices gives some mathematical details, thus saving constant reference to other books.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

1. Introduction

As reliability has become a more formalized activity, the body of knowledge which it includes has expanded and become compartmentalized. The particular subset called "parameter variations analysis" is described literally by the title. It is the analysis of the changes in parameter values (figures of merit, performance, cost, reliability, electronic properties, mechanical properties, etc.) when some of them are varied, some are constrained, and the rest are allowed to change as necessary. It may take place at any systems level from raw materials to supersystems. The tools are physical models and mathematical equations.

This report is a tutorial exposition on the state-of-the-art of parameter variation analysis. The emphasis is heavily on the practical nature of parameter variations analysis and its usefulness. This report is intended to give the engineer an overall view of the situation and to give him practical suggestions about utilizing parameter variations analysis. It presumes the reader has knowledge at least equivalent to a B.S. in Engineering although much of it may have been forgotten. It is not a cookbook in the ordinary sense of the term, even though there are formulas in it. This report is a vehicle to aid understanding and is organized with that view in mind.

Chapter 2 discusses in detail some of the concepts which are important in parameter variations analysis. Chapter 3 is a fairly brief listing of the uses of physical models. Chapter 4 is concerned with the creation of mathematical models of systems. The techniques for analyzing these models are expounded at length in Chapter 5 which is subdivided according to the importance of probabilistic considerations in the models. It emphasizes the essential fewness of the techniques for analysis. Chapter 6 points out the diversity of uses of mathematical models as opposed to the techniques for the analysis given in Chapter 5. Chapter 7 is a brief discussion of the sources and uses of data for these equations. The report is concluded with the usual summary and conclusions.

The ideas of a conceptual model and of models versus reality are important to the point of view pervading much of this report. Therefore Sec. 2.1 (models) should be read carefully.

2. Concepts

There are several concepts which are used in this volume, or in connection with this subject, and about which there is some confusion. It is the purpose of this chapter to discuss those concepts and give them fruitful meanings.

2.1 Model

The idea of a conceptual model is adapted from the idea of a physical model such as a model car or the model of a building. In a physical model, the characteristics of importance to us are reproduced quite well. In a model car these might be proportions, shape, and color. The ones of little or no importance are not usually reproduced at all; e.g., there might be no motive power and the tires may not be pneumatic. The "inbetweens" receive indifferent treatment; e.g., the windows may be transparent and the presence of seats inside may be inconsequential. The physical model is an abstracting of something important from the real world; it is an imitation.

A conceptual model is analogous to a physical model. Since everything in the universe probably affects everything else to some degree, however slightly, any exact treatment would be hopelessly complicated. Therefore we decide how we will look at the situation and make a set of assumptions (both explicit and implicit) about what we will ignore and what we will include in our conceptual model. It is usual to state only a few things that are being ignored and to make the blanket assumption that everything else which is not explicitly mentioned is also to be ignored. By its very nature, a conceptual model is incomplete: it ignores some things and describes other things in an approximate fashion.

After having made a set of assumptions for our conceptual model, we operate on those assumptions with mathematics and logic; we analyze them by any means at our disposal. The assumptions together with the current results of the analysis are our model. Unpleasant situations sometimes occur while developing the logical implications of a set of assumptions - we do not like the results for one reason or another - they do not seem to fit, they appear to be inconsistent with our beliefs, etc. Under these circumstances we have two rational choices:

(1) Change our beliefs about the way the world is if we are convinced that the set of assumptions are very realistic, or

(2) Go back and modify the assumptions so that their logical implications do, in fact, fit our beliefs about the world.

The creation of a conceptual model is a circular, often haphazard, process wherein ideas come from everywhere, get analyzed, tested, compared, junked, and accepted. Some good ideas usually filter through the process.

The completely logical structure of a conceptual model is developed after an idea is successful. Some of the ramifications are so complex that it takes much calculation to find out what they are.* Sometimes we refer to one equation or curve as the model, but this is just speaking loosely.

If a model fits the real world well enough for our purposes at the moment, it is an adequate model for the moment. Adequacy depends not only on the model and the world themselves, but on our needs and desires - not to mention our ability to compare the model with the world. Thus, models are not right or wrong but are only more or less adequate. Of course, some models are so woefully inadequate for anything that we class them as wrong. Others are so generally adequate that we feel they correspond very closely to reality.

In this latter case, however, it is important to distinguish between a definition and a model; the reason some "models" do so well is that they are, in fact, definitions of one of the quantities or concepts involved. For example, "An unbiased coin toss will have a 50-50 chance of heads or tails" is not as assertion about the world, but a definition of unbiased-coin-toss." If it does not come out 50-50, we do not change our ideas about what unbiased coin tosses do, we search for the bias in the coin toss.

An engineering model is often mathematical in nature and the same formalism will describe several different situations. For example, the equations which describe resistance-inductance-capacitance networks will also describe mass-spring-dashpot systems. Furthermore there is more than one analogy between the two that can be made. It is important to keep the distinction between the mathematics itself (which is quite general and completely impersonal) and what we have it represent in an engineering sense. The mathematics never lie, but often they do not apply.

In some cases we can not or will not write down equations but rather we discuss the system at the level of phenomena. This is called a qualitative (phenomenological) model.

The term "probabilistic model" appears in the literature (but not here). It is generally a special case of a mathematical model wherein the relationships are between probabilities or between random variables.

We never analyze the real world, we can only analyze a conceptual model of the real world.

2.2 Randomness

Randomness is a rather basic concept; it is difficult to define without being directly circular. The best we can do is talk about it enough to make sure that we

---

* Thus the science of simulation.

all have similar feelings for it. First, we can say that if the mathematical theory of probability applies to the events or variables then those events or variables are random. Next we can say, from an engineering viewpoint, that if the uncertainty in the events or variables is appreciable then they are random. There is no distinction between cause-and-effect and randomness. Outside of concepts where quantum-mechanics is necessary - and this never happens in Reliability--a cause-and-effect relationship is presumed to hold for all physical events. But often we do not wish to go into details, or virtually cannot as in statistical mechanics, and so we use a statistical description of what happens. Even though a parameter may have a specific value we may assign a number (a probability) to our degree of belief about the possible values of that parameter. If we are applying probability in connection with something, it is, by definition, a random event or a random variable, regardless of any other considerations.

It should be emphasized that the theory of probability is not limited in its application to the relative frequencies of events any more than a particular differential equation is limited to an electronic circuit vs. a mechanical system. One of the early applications of the theory of probability, and one of increasing activity today, is to use probability as a measure of degree-of-belief about events or parameters. This is sometimes called a Bayesian approach. Some articles would have you believe you must adopt either the degree-of-belief ("subjective") or the relative-frequency ("objective") approach to probability and that you cannot use both. This is not so, although it is wise to be extremely cautious about mixing them; you can use either one when it suits your purpose. Even though the mathematical theory of probability is quite useful when discussing random events and variables, it is sometimes very difficult, if not virtually impossible, to associate a particular probability with certain kinds of knowledge.

There is confusion in the literature about the meaning of "pure chance," "purely at random," and similar phrases. They are often used when the hazard rate (for time to failure) is constant or when the number of failures in a fixed time interval has a Poisson distribution. This is poor practice since it tries to impart a degree to randomness. But one would not wish to say that a Normal (Gaussian) variable was more random than another because its variance (or standard deviation) was larger. It is better to compare descriptors such as variance or coefficient of variation, or to describe the probability density function (pdf) directly; e.g., "the pdf is uniform, i.e. constant, over the range $0 < x < a$." For another example, rather than saying the distribution of heads and tails on a tossed coin is given by pure chance, it would be better to say that the probability of heads or tails for this kind of toss is one-half.

There is a very close link here between randomness and uncertainty. Events are not random in themselves, but only relative to our needs, desires, and ability to measure and predict. For example, suppose some rods are going to be hammered into the ground and used as markers. If their lengths are all 18" $\pm$ 1," we might say they are all the same, the length is not uncertain; on the other hand, if these rods are to mate with a cast metal part, their length is quite uncertain and we would treat the length as a random variable.

2.3 Parameter

In mathematics one sometimes distinguishes between parameters and variables. Parameters are usually held constant during the course of an analysis and variables are allowed to change. This naming implies some knowledge of how the values of each will behave. When creating a model for physical systems such prior knowledge often does not exist or is completely arbitrary. For example, the dimensions of a bar or resistance of a resistor might be fixed or changeable depending on the nature of the analysis.

In order to avoid the implications of fixed and changeable, only the name parameter will be used for describing all quantities to which are assigned an algebraic designation or letter. That is, it includes both the parameters and the variables of mathematics - any attribute of a system, part, component, etc., to which numerical values can be assigned, and to which no permanent numerical value is given, is called a parameter. The term is intended to be very general: it covers inputs, outputs, properties of materials, environmental descriptions, forces, deflections, strengths, figures-of-merit, etc. On occasion in discussing statistical distributions it is convenient to use the word parameter and variable as is customary in statistics and it is done in the text where it will avoid confusion. For example, the mean (μ) and variance (σ^2) of a Gaussian variable are referred to as parameters. It is customary to denote the parameters in statistical distributions by Greek letters.

This procedure, adopted here, is completely arbitrary, but is less cumbersome than using both words (parameter/variable) where the meaning would not otherwise be clear.

2.4 Reversible vs. Nonreversible

It is convenient in engineering to be able to use these terms without implying the thermodynamic definitions. A process is reversible if the system can and may be brought back to its original state by traversing backwards through the subsequent states; otherwise it is nonreversible. Implicit in this definition are several concepts:

(1) The system is described by listing its important parameters. Only these need

be brought back to their original values. There will be some parameters which, while not important, are not ignorable either. They should be brought back closely enough to their original values.

(2) There is some feasibility and desirability implied besides possibility.

The terms are a matter of degree, not a black vs. white situation; what is reversible in one context may be nonreversible in another.

An important case is one where a dependent variable is a function of several independent variables, none of which is time. If the independent variables represent things you do to an element, such as change its temperature, and the dependent variable represents a measurement that is made, such as resistance or length, the process is reversible. The word function is used in its strict sense, viz., it is single-valued and defined everywhere in the range.

Often there is hysteresis present. If it is small enough, the process is reversible. If it is too large to ignore, then the process is not reversible, although the term nonreversible may be too harsh for it. Do not be confused by the necessity of applying a label to the process - just describe it as well as you can.

2.5 Drift and Degradation

These terms are applied to nonreversible processes only, but the changes need not be monotonic. If most of the system parameters and all the external, ambient conditions are at their original values* and a few are not, those few are said to have drifted. (This assumes there is no reversible relationship between those few.) There is usually some implication that the process was a slow one (compared to the time scale of concern).

If the drift was bad, and this is a value judgment, it is called degradation. The judgment on a particular drift may well be different at different occasions. Thus a 1% change in resistance or in deflection of a beam is not good/bad in itself, but only in relation to other circumstances.

2.6 Performance

This word has several kinds of uses. It generally refers to some figure-of-merit (FOM) of an element or item. A system often has several measures of performance and just as often, there must be tradeoffs between them. This term will rarely be used in this volume. Parameter is used to be more general, FOM to be more specific.

---

* This refers only to the ones we are measuring or controlling, and even then, only to the degree we are doing it. Thus if ambient temperature is ignored, we would call the results of a temperature change "drift".

2.7 Item or Element

The terms "item" or "element" are used in a general sense since they have not yet been preempted by anyone to refer to a specific size or group. It may be a part, component, subsystem, system, or collection of systems. It may be as large as SAGE or as small as an integrated diode.

2.8 Figure-of-Merit

A figure-of-merit (FOM) is just what the name literally implies. It is a figure by means of which the merit of an item may be determined and by which items may be compared and ranked. The FOM is generally presumed to be directly related to values upon which decisions are made. Reliability, signal-to-noise ratio, strength-to-weight ratio, and gain-bandwidth are all examples of FOM's. An FOM is a parameter and equations can be written showing its relationship to system makeup and to other FOM's. Many measures of performance can as well be called FOM's since the merit of systems is judged by performance among other things.

The FOM need not be dimensionless although many are. Some are normalized to be perfect for FOM = 1 and worthless at FOM = 0. This is all a matter of taste and convenience.

2.9 Statistical Ignorance

While an engineer may refer loosely to his knowledge on a subject as "complete ignorance" and be reasonably satisified that he knows what he means, it is difficult to quantify this state of complete ignorance. In fact, almost any attempt to quantify it has logical implications which appear to be contradictory. Statisticians have been concerned and argumentative about quantifying ignorance since the beginnings of formalized probability theory. It is wise to remember that we are dealing in abstractions from the real world, i.e., conceptual models. If any of the logical implications of a set of assumptions appear to contradict what we think we know, we have several choices, among which are:

(1) to change our idea of what we think we know; i.e., we feel we were mistaken,

(2) to go back and change some of the assumptions so that what we derive from them is consistent with our observations of the world.

In the early development of a conceptual model this latter course is most often followed.

Another situation of relevance occurs when the mathematical formalisms used to represent different models are the same. While in some ways we can then say the models are equivalent, we must be careful not to call the models the same because they are, in fact, talking about different things. For example, the resistance-inductance-capacitance equations can turn out to be exactly the same* as some mass-spring-dashpot equations, but the models are of different things.

---

* There is not a unique one-to-one correspondence between the two; several analogies exist.

In the field of probability the same mathematical theory is used to describe at least two different situations which, while related, are certainly not the same things. The first is relative frequencies of events; gambling theory is based upon it and much of the theory of probability is shown to be quite applicable to this situation. Be careful though about nominal vs. actual situations; loaded dice and fixed one-arm bandits are not entirely unknown in the real world.

The second use of the mathematical theory of probability is to describe what is known as subjective probability which refers to our feelings on matters and to the the way in which we might place bets.* In cases where there is some overlap of these two, the prudent man will have his subjective probabilities close to the relative frequencies. The mathematical theory of probability does not have to refer to one or the other of these two any more than a second-order linear differential-equation has to refer to either the electrical or the mechanical system. It can equally well describe both or neither.

One can avoid philosophic arguments about what probability really is by asserting that he is using the mathematical theory of probability to represent a certain situation and that it is believed to represent it adequately in the circumstances.

In the relative frequency model it is considered irrational in many situations to assign a prior frequency distribution to a parameter.** For example, the true mean strength of a set of columns is what it is regardless of our state of knowledge about it. On the other hand, my feelings about what that true mean strength is may well have a probability distribution. If the subjective probability, which this latter is called, is being used and represented by the mathematical notation for probability, care should be used not to confuse it with relative frequency especially part-way through a problem.

If a parameter is considered to have a probability distribution it often seems desirable to represent the situation wherein the engineer says, "I don't know anything about it; I am completely ignorant." Very often people have said, "In that event, let me assume that all values of the parameter are equally likely." There is no intention of discussing the moral uprightness of those who do so; but there are technical problems with this approach. For example, consider the single parameter in $pdf(t) = \lambda\exp(-\lambda t)$.*** Now suppose that:

---

* The fact that people are not always rational and have other peculiarities, means that there are situations to which a rational conceptual model may not apply.

** There are some situations in the relative frequency model wherein the parameter can be legitimately considered a random variable, but it is not the intent here to discuss when this happens.

*** This is the exponential distribution for time to failure.

(1) Probability refers to my degree of belief (i.e., it is "subjective") about λ.

(2) I wish to assert that I am completely ignorant about λ prior to running an experiment (but after giving it some thought).

(3) I choose the uniform distribution to describe this ignorance.

Several difficulties arise as the logical implications are considered:

(a) Obviously λ is limited to non-negative real numbers,

(b) Since the interval is semi-infinite, the pdf would not be defined. At this point a quasi-pdf could be introduced as in Ref. 1. For an engineer, it is more reasonable to pick an upper limit (λ_2) above which $\text{pdf}(\lambda > \lambda_2) = 0$ since no one really thinks a λ of, say, $10^{1000}\,\text{sec}^{-1}$ is realistic for any equipment worth considering.

(c) Likewise no one really expects this equipment to have a $\lambda < 10^{-1000}\,\text{yr}^{-1}$ (or some other small number) in the field. Therefore it is convenient to pick a λ_1 such that $\text{pdf}(\lambda < \lambda_1) = 0$.

Now I no longer have the original assumptions. Instead, the third assumption is modified to:

(3') I choose the uniform distribution over the interval λ_1 to λ_2 to describe this ignorance, viz.,

$$\text{pdf}(\lambda) = \begin{cases} 0, & \lambda < \lambda_1 \\ a, & 0 < \lambda_1 < \lambda < \lambda_2 \\ 0, & \lambda > \lambda_2. \end{cases}$$

We can now reach the next difficulty:

(d) Consider $\text{pdf}(1/\lambda)$. It can be shown by probability theory that

$$\text{pdf}(1/\lambda) = \begin{cases} a\lambda^2, & \lambda_1 < \lambda < \lambda_2 \\ 0, & \text{elsewhere}. \end{cases}$$

Obviously if (3') describes complete ignorance, then (d) does not, because the $\text{pdf}(1/\lambda)$ is not a constant in the non-zero region.

The easy way out of this problem (and the best way for engineers) is to eliminate the term complete-ignorance and say that "$\text{pdf}(\lambda) = a$" represents my state of beliefs rather than complete ignorance. If I do not mind my state of belief for $1/\lambda$ being what it turns out to be, then I am all right and there are no contradictions involved. In this particular situation some have responded by asserting that $\text{pdf}(\ln\lambda) = \text{constant}$ (over some range) represents complete ignorance, but functions of λ other than $\ln\lambda$ can show the contradiction.

Generally, the engineer is not as ignorant as he claims to be in connection with his beliefs about the parameters of a distribution. In fact, engineers have been known to make the statement, "I am designing this equipment with a $\lambda = 1/3\lambda_{spec}$ in order to pass the test." The engineer would not then wish to describe his subjective probability for λ with a uniform pdf, but rather one that had at least a peak in it near $\lambda_{spec}/3$. The peak might be rather broad, but nevertheless it would be there. A way to evaluate these probabilities is to guess how you would like to bet money, write it down, evaluate some of the logical implications of that decision, and see if you are willing to live with them.* If not, go back and modify the original choice, etc., until you arrive at a model which you are willing to tolerate at least.

In general, no one really insists on complete ignorance (about anything he has even heard of) to the extent that he will not back down from, "I wish to assign to my feelings on the matter equal probability density anywhere from minus infinity to plus infinity or, in the case of a number which is known to be non-negative, from zero to infinity." When really pressed on the matter it turns out that his non-ignorance will take shape. For example, he may be quite sure that a hazard rate (λ) has $1/\mu sec > \lambda > 1/1000yrs$; now he is talking about knowledge and belief rather than complete ignorance.

When calculations are to be made from this prior distribution by using Bayes' formula** it is wise to pick a function that is as tractable as possible. This is not usually too difficult since many of the tractable functions have enough arbitrary parameters and the general shape to give the loose fit desired to the rough prior probabilities.

In the paradox of complete ignorance of parameters there is a striking resemblance to the paradox in the method of maximum entropy (this method was developed by E. T. Jaynes and appears sporadically in the literature). It is easy enough to

---

* In trying to create a pdf for subjective prior probability and to evaluate the tentative curves suppose that your malevolent brother-in-law were to use that curve to establish odds, choose his own side, and bet money. Would you then wish to change the curve so as not to lose your shirt and more to him? If so, change the curve to one wherein you are willing to let someone try to take advantage of you by using it to calculate odds for wagers and then by choosing his own side of the wager. Finally you have really gotten your feelings on the matter quantified.

** There is no controversy about Bayes' formula itself, there is only controversy about the kinds of things people represent by mathematical probabilities when they are using Bayes' formula. The formula is $P(A|B)P(B) = P(B|A)P(A)$; see Sec. 2.10 for notation.

use formally, but the results in practice are quite sensitive to the choice of parameter. Information (or its opposite, called entropy) is defined with respect to a certain parameter and has the same limitations as complete ignorance does.

In summary, the terms "ignorance" or "complete ignorance" are not accurately descriptive and are likely to lead to contradictions when made quantitative. Therefore it is better to try to describe the state of knowledge or belief by an appropriate prior* distribution.

2.10 Statistical Independence and Linear Correlation

Independence and correlation are very technical terms in the statistical literature. The confusion between them generally arises because even though they are defined very precisely, the terms themselves do not readily convey their meanings to engineers. For that reason, in these volumes on PRACTICAL RELIABILITY, they are generally written as "statistical independence" and "linear correlation" which help to convey the technical meanings.

There are two definitions for statistical independence, one for discrete events and the other for continuous probability density (pdf). Consider first the case of discrete events. The notation adopted is as follows.**

A, B, are identifiable events

$\overline{A} \equiv$ not A.

$P(A) \equiv$ the probability that A occurs.

$AB \equiv BA \equiv$ the joint event of both A and B.

$P(A|B) \equiv P(AB)/P(B) \equiv$ probability of A when it is known that B did happen.

We are now ready to define statistical independence for discrete events:

If, and only if, A and B are statistically independent,

$$P(AB) = P(A)\,P(B), \qquad (1a)$$

$$P(A) = P(A|B) = P(A|\overline{B}), \qquad (1b,c)$$

$$P(B) = P(B|A) = P(B|\overline{A}). \qquad (1d,e)$$

Any one of these five equations can be used as the definition of statistical independence; the others then follow from the laws of probability. Equation (1a) is most often used in the literature as the definition of statistical independence although it is easier for many engineers to visualize (1b,c) or (1d,e) as a definition. It should be noted also that if A is statistically independent of B, B must be statistically independent of A and so it is stated only that A and B are statistically

---

* Prior means prior to the next experiment, not prior to careful thought about it.

** A summary of probability theory is given in an Appendix of Vol. IV - Prediction. The symbol $\equiv$ means "is defined as" or "is identical to".

independent. In the statistical literature, the modifier "statistically" is rarely, if ever, used. But in engineering literature it should always be added (where it is meant) so that there is no confusion with physical independence (physical independence may be loosely thought of as lack of cause and effect).

If there are more than two events, the requirement for complete statsitical independence of the events is usually written as $P(AB...C) = P(A) \cdot P(B) \cdots P(C)$. Subsets of the events may be statistically independent without all of the events being so; but if there is complete statistical independence for all events, then the contents of any and all subsets are statistically independent.

Now consider the case where the parameters are continuous variables and have continuous pdf's. Then the definition of statistical independence is: the continuous random variables, $x_1, x_2, \ldots, x_n$, are statistically independent if, and only if,

$$\text{pdf}(x_1, x_2, \ldots, x_n) = \text{pdf}(x_1) \cdot \text{pdf}(x_2) \cdots \text{pdf}(x_n). \qquad (2)$$

That is, the joint pdf is the product of the individual pdf's. If there are more than two parameters, they can be pairwise statistically independent without being all statistically independent. If they are all statistically independent any and all subsets of the parameters are statistically independent.

It is oftentimes difficult to tell from physical reasoning whether two parameters are statistically independent or not. Making the decision may require a fair amount of study and effort. Statistical dependence is rarely discussed as such but events or parameters are statistically dependent if they are not statistically independent. There is no definition of the degree of statistical dependence,* although one could probably define a complete statistical dependence as a situation wherein the conditional probabilities are either 0 or 1.

Two parameters can easily be statistically dependent by virtue of their relationship to a third parameter, the relationship between the first two obviously not being one of cause and effect. Such may be the case, for example, between a telephone system's working poorly and many people carrying umbrellas - both can be due to wet weather.

Correlation is an ambiguous word to the engineer because it seems to mean the same thing as statistical dependence. However, it is synonomous with linear correlation as far as statistics is concerned. The linear correlation coefficient (ρ) is defined

---

* If there is linear correlation, the correlation coefficients might be used for this purpose.

by the equation,

$$\rho\sigma_1\sigma_2 = \text{covariance } (x_1, x_2)$$

where the σ_1 and σ_2 are the standard deviations of x_1 and x_2,

$$\text{covariance } (x_1, x_2) \equiv \int\int (x_1 - \mu_1)(x_2 - \mu_2) \text{ pdf } (x_1, x_2)\, dx_1\, dx_2,$$

$$\text{variance } (x_i) \equiv \sigma_i^2 \equiv \int (x_i - \mu_i)^2 \text{ pdf } (x_i)\, dx_i.$$

Generalized product moments can be defined as in Ref. 2.

There are well known examples wherein the linear correlation coefficient is zero (the variables are said to be uncorrelated) yet there is a direct functional relationship between the two. A circle with its center at the origin is a good example. The important thing to remember is that "correlated" means "linearly correlated" and that two variables may well be linearly uncorrelated yet statistically dependent. In the literature this statement would read, "the variables may well be uncorrelated yet dependent." The latter can be most confusing to those not familiar with the ellipsis being used.

2.11 Statistical Estimates*

All statistical estimates are, in their essence, point estimates; any interval estimate consists of two point-estimates, one for each end point of the interval. For example, one can use the sample mean (a statistic**) to estimate the true mean.

Then proceed as follows.

(1) From the pdf of the estimate of the mean (this estimate, remember, is a statistic) calculate two end points for a particular confidence interval. Call them EPU and EPL (for End Point Upper and End Point Lower).

(2) Each of these two end points - EPL and EPU - is a statistic since it it calculated from sample data. For simplicity consider only one of them, e.g., EPL. It has a pdf since it is a statistic; so calculate two end points for a confidence interval for EPL. Call these two end points EPL (EPL) and EPU (EPL) for the lower and upper end points respectively of the confidence interval for EPL.

---

* A summary of fundamental concepts for statistical estimation is given in the Appendix of Vol. III - Testing.

** A statistic is any result obtained by manipulating the data. Any statistic which is derived in a known way for a random sample from a known population (regardless of the population size or sample size) has a pdf. It may not be known nor expressable in a tractable form but it is there nevertheless. Just suppose there are a very large number of replicas of the population and take, in a fixed way, a random sample from each replica. Calculate the particular statistic for each sample. In this example, calculate the sample mean--this is the statistic being used to estimate the population mean. Now plot its distribution and there you have it--the pdf of the statistic under consideration.

(3) Each of these two end points - EPL(EPL) and EPU(EPL) - is a statistic since it is calculated from sample data. For simplicity consider only one of them, e.g., EPU(EPL). It has a pdf since it is a statistic; so calculate two end points for a confidence interval for EPU(EPL).

(4) Note that (2) and (3) are alike except for the names of the end points. One can keep this up until he tires of it.

This procedure illustrates the fact that all estimates are, in their essence, point estimates. As a matter of fact, it is virtually unheard of for anyone to go beyond the first confidence interval. But many engineers have suggested confidence on confidence...in derision because they objected in the first place to the unfamiliar concept of confidence.

We shall limit the meaning of the word "estimate" here to a number derived from the data (i.e., a statistic); but any number is an estimate no matter how poorly derived or wildly conceived. It may be a particularly poor or misleading estimate; it may have absolutely none of the good properties and all of the worst properties ever associated with estimates; but it is nevertheless an estimate. Statisticians describe properties of estimates by various terms such as efficient, sufficient, unbiased, maximum likelihood, least squares, and minimum variance*. In some cases a particular estimate can have many of these properties at once. In other cases the estimate cannot or does not have them all. Most of these properties are considered good ones; e.g., an unbiased estimator is better than a biased one (other things being equal). But the tradeoffs between them are not clearly defined. Generally, an engineer will take the best he can get without making tradeoffs. Then the uncertainty in the estimate is likely to be larger than any modifications to be made by the tradeoffs. Thus it becomes a matter of indifference, of personal preference, and of tractability as to which ones are used.

* These terms are not mutually exclusive.

3. Physical Models and Their Uses

Physical models are variously called breadboards, mockups, etc.; it is not difficult to put prototypes and preproduction samples in this class also. It is convenient sometimes to include even the hardware itself as a special case of a physical model.

3.1 Design Assistance

With a physical model, parameters can be changed by substituting or varying the elements. There are several ways of doing this:

(1) Consider one particular element. Substitute other similar ones for it, one at a time, and measure all the appropriate figures-of-merit each time (e.g., gain of an amplifier, audible noise level of an hydraulic system). The similar elements may have unintentional differences or may be selected for their differences. Watch out for correlations between parameters on the same element. Repeat this for other elements. For a more efficient method, see (2) below.

(2) The above process can often be made more efficient if a statistical design is used for substituting several parts at once. Some knowledge of the system is required over and above that for (1) above if the increased efficiency is to be realized, but interaction effects can be found this way that would likely be missed in (1). There are factorial, partial factorial, Greek squares, Latin squares, nested designs, ad nauseum. This is called Experimental Design and the literature is full of explanations and detailed plans.* But beware of simplified explanations in the engineering literature--they are often misleading. Professional help from statisticians is virtually a must, but be sure you understand all the assumptions and approximations in the experimental program layout and in the analysis. You need not understand the details of the design or the analysis, just understand what the statistician is assuming your system is like. If you abdicate your responsibility, you may find that you have gotten worthless results--correct from a statistics point of view, but worthless from an engineering standpoint.

(3) As a special case of the above, several replications of the system may be made instead of changing the elements in one system. Thus for example, 10 to 20 electronic cordwood modules, with appropriately chosen element values could be made and tested. The results would be analyzed to show how the figures-of-merit depend on the element parameters. Again, correlations between several parameters of a given element must be watched; they complicate the analysis.

---

* The Appendix of Vol. III--Testing in this series also discusses experimental design.

(4) On a breadboard or mockup, it is sometimes possible to replace some of the elements with similar ones which are adjustable. For example, a variable speed motor might replace a constant speed one and a rheostat might replace a resistor. Schemes have been developed for simulating decay of gain in a transistor, etc. A difficulty with this method is that the substituted element in itself is only a partially adequate model for the actual element. In some ways it won't behave the same way, e.g., the adjustable speed motor may have a different rotor inertia and torque-speed curve and the rheostat has a different high frequency response. This method can be especially useful in early design.

(5) Some commercial devices are available for automatically switching a "high" and a "low" valued element in each of many positions for electronic breadboarded circuits. This can be useful in worstcase design. The difficulties in (4) are still present however.

Even though none of the above methods is always suitable and all have some bad points, all are sometimes useful. Where a method has been systematically carried out, the results can be profitably codified into curves or equations. In some cases, a statistical analysis of the results can be used to fit a linear, quadratic or specialized surface to the data and to get an idea of the uncertainties involved in prediction using these equations.

Care should be used in evaluating the results of tests on the physical model. Very often, electrical characteristics only are measured while thermal and mechanical considerations are ignored. In that event, special models must be constructed to evaluate the thermal and mechanical properties. On occasion when this has been done, the electrical behavior of the more apt model has changed from the more simple-minded one and troubles begin. Other properties such as optical and radiation may be important; then the model must be adequate for these as well.

One reason for testing a physical model is to check out a mathematical model. In general, this measures the adequacy of the mathematical model rather than the physical one although cases may arise wherein the physical model had unrealized inadequacies. The blind insistence on one of these being true to the exclusion of the other can cause lost time and money plus many ruffled tempers.

3.2 Production Assistance

In addition to being useful for design purposes, physical models are valuable in determining the effects of variations in manufacturing processes, before the system is finally put into production. Ordinarily the variations analysis for manufacturing purposes is much simpler than it is for design assistance but it

nevertheless should be investigated for high reliability applications. Mockups and other physical models are one of the tools. If the quality of particular parts is extremely sensitive to something in the manufacturing process, a design revision may be called for. At other times just a change in the type of manufacturing process may be necessary. For example, the method of heat treating a steel part may have to be changed.

4. Creating Mathematical Models

Before one can analyze a mathematical model of a system, one must first create it. This is usually done by subdividing the system into subsystems that can be readily handled. Very little new has been developed in the past decade or two for the creation of models. But there have been tremendous advances in the ease with which it can be done particularly by the use of computers.

4.1 Subdividing the System

The system must be subdivided for analysis into elements which are small enough to be handled both by the means at ones disposal and in the manner which makes it most convenient to get answers. In some cases the entire system will be broken down into elements for which equations describing the performance (or FOM)* of each element are known by inspection. Such elements as resistors, capacitors, simple tuned circuits, and class A amplifiers often have equivalent circuits which are quite adequate.

There is no arbitrary rule on how small an element must be when subdividing a system. It must be merely small enough so that it can be handled. For example, a simple parallel tuned circuit consisting of a parallel resistor, inductor, and capacitor could be treated as an element itself if the equation were available; it could be broken down into two elements, e.g., the resistor and the LC combination; or it could be subdivided into three elements. What is done depends on the user's knowledge and the adequacy of his references, or on his computer program.

Remember that an equivalent circuit for a device is essentially a conceptual model, and the equations for it are implied (in the mathematical or logical sense) by the equivalent circuit. In fact, the symbols we use have two separate functions:

(1) to represent the physical part, or

(2) to represent the equation for the part's operation.

Most equivalent circuits are valid only over a range of the variables or parameters and care must be exercised in using them that the parameters do not exceed the allowed range. For example the small-signal equivalent-circuit for transistors is not generally valid for an oscillator circuit. Mechanical springs may have a distributed mass which becomes important at times and their nonlinearity may be important beyond a certain extension.

There will be many occasions where a nice, neat equivalent circuit for an element does not exist and it is not convenient to break it down into smaller elements. In that case, it is customary to go into the laboratory and measure the performance of the element. Sometimes there is a form that the results are expected

---

* Figure-of-merit.

to take* and this form is known analytically. If the results appear to behave in the predicted manner adequately enough, the unknown parameters in the form (equations) can be evaluated by means of the data. In other cases the results are merely tabulated or a curve is fitted by the most likely looking interpolative functions,** without regard to any other kind of analysis.

A distinction is sometimes made betwen theoretical and empirical equations. It should be remembered that all equations in use to describe physical elements are empirical in their essence. A theoretical equation is one that someone derived empirically a long time ago and has been honored by the passage of time and at least some success. Many are equations which describe the behavior of elements such as springs and resistors, or processes such as viscous flow. A sophisticated name for them is constitutive equations (as opposed to the conservation equations).

Most of these constitutive equations have one parameter which describes a property of the element, fluid, etc., e.g., resistance, spring constant, viscosity, and diffusivity. It is very convenient to look upon these equations as defining that one parameter. There is then no question of the equation's being true, because true/false doesn't apply to definitions.

Often these equations are called laws but, obviously, it can't be a law and a definition at the same time. It is most convenient, even if not historically accurate, to call the equation a definition; then the law describes the parameter. Take Ohm's Law for example: resistance is defined by $R \equiv E/I$ (the constitutive equation). Ohm's Law then states that the R, so defined, is constant*** (independent of V or I) for many materials over wide, usable ranges of the variables V and I. Fick's Laws for diffusion, Hooke's Law for springs, and Newton's Law for viscosity, as further examples, can be treated in the same way. Of course, for all these examples, materials are well known for which the Laws (of constancy) do not hold.****

---

* This form is often derived from a very simple-minded conceptual model of the element.

** Finite portions of infinite series for orthogonal polynomials, for a simple power series, or for trigonometric functions are often used. The choice depends on ease of calculating the coefficients, the criteria for goodness of fit, the kind of function, and the ease of making the calculations from the equations.

*** The range over which the parameter is constant depends on the precision involved. A spring may have a constant k, within 1%, over the range from zero to maximum compression, but is constant over only 1/10 that range to within 0.03%.

**** Very often, differential parameters are defined in these cases, e.g., $\tilde{R} \equiv dE/dI$ is called the ac resistance and is very useful sometimes.

The amount of detail necessary in developing the mathematical model depends on the available facilities. For example, some programs for digital computer analysis have equivalent circuits for some of the more common electrical parts such as resistors and capacitors. In these circumstances, the full equations for the system are never written down but are contained implicitly in the computer program. They nevertheless are implied by the computer program itself and by the data inputs to the computer. The fact that the final equations are only implicit rather than explicit in no way gets around the fact that they do determine the results.

Engineering handbooks (mechanical, electrical, electronic, etc.) are a most valuable source of existing conceptual models for elements.

4.2 Traditional Engineering Analysis

It should be pointed out that our methods of analysis ordinarily have separate models for the thermal, electrical, and mechanical behavior of a system. There may be overlap in some areas but usually the models have been developed so as to keep these overlap areas to a minimum. During the development of a system model, it helps to keep in mind the admonition that the order we see in the world or a system is one imposed by us, not one discovered by us. Thus, there may be several different methods of analysis for the same system which are equally fruitful.

Everything in the world affects everything else in the world all of the time but most of these effects are negligible for any particular system. The more precise and exacting is the analysis of a system, the more must some effects be considered which were previously neglected. For example, systems which are concerned with length accuracies on the order of 10^{-4}% must be concerned with temperature changes induced by the presence of a person. Systems which deal with fractions of μVdc must take into account small emf's due to chemical and thermal differences which are always otherwise neglected (in some precise dc potentiometric applications the kind of solder used on the joint is most important). It is a matter of judgement and total resources as to what effects should be included in a system analysis. Quite often special experiments are run, (sometimes simple and sometimes very complex) to determine whether effects are small enough to be neglected or large enough so that they must be included.

In this volume we are interested in how the system behavior is modified as the parameters used to describe the elements change somewhat. It is therefore important that the equations used contain explicitly the parameters which we are interested in varying. Most conceptual models traditionally deal with parameters in the usual range of operation of the system. Some recent work has been done wherein the model

describes the system in the region of failure, near-failure, and abnormal operation. In parameter variations analysis it is a good idea to write down for each equation the range of parameters over which that equation is expected to be valid.*

When one of the specialized digital computer programs** is used for the analysis, it is important that the engineer know the basis on which the program was developed. The degree to which he carries this is governed by how pressed he is by other matters, the resources available to him, how accurate his analysis must be, and how lucky he feels. For example, the electronic equivalent circuits which have been assumed for elements should be satisfactory to the engineer who is using the program but he may take them on faith if he's in a hurry or has had good luck with them before. In many cases, the parameters required by such a program as inputs are not the parameters in which the engineer is interested nor are they parameters which the manufacturer ordinarily specifies. Programs do exist for transposing between the two but they too contain assumptions which should be explicitly identified if high reliability is essential.

Nontraditional conceptual models for systems are being put in explicit form. For example the behavior of many models is determined by the location of poles and zeros in a Laplace transform. In some cases, techniques are available for analyzing the system directly from the knowledge of the coordinates of the poles and zeros. If parameter variations of the parts can be transformed into variations of pole and zero locations, then parameter variations analysis can be accomplished.

It is important to remember that the techniques for parameter variations analysis are not limited to such things as deflection of a beam or output of an electronic circuit, but may be cost effectiveness, delivery time, or anything for which an equation may be written. The analytic techniques are the same.

4.3 Automated Procedures

Some typical computer programs which are suitable for use in analyzing systems are given and discussed more in Vol. II--Computation (Chapter 4) of this series. The computer program is the framework of the model for the analysis of the system.

---

* This should always be done in any exacting analysis of a system anyway; it would save many mistakes from being made and serves to point up places where engineering knowledge of the system is deficient. This process in itself will usually require somewhat more understanding of the system than was needed to write down the equations in the first place.

** Typical ones are identified in Vol. II--Computation (Chapter 4) of this series.

It specifies the kinds of elements that are allowed and the equations used to describe them (constitutive equations). The computer does not create the model (in the sense in which this volume uses the word) for the system since the framework is already in the program. Making an abstraction of the actual system, in the form which can be used by a particular program is the job of the engineer, and in this sense he is creating the model of the system within the constraints of the program. The computer then performs the logical steps (including arithmetic) to give the answers.

Some computer programs are much more amenable to parameter variations analysis than are others, although many which heretofore were not suitable are being modified in the appropriate direction. Computer programs have the great advantages of simplicity, ease of use, short time of analysis, convenient printout of the results, relative freedom from errors of calculation, comprehensiveness, and no forgetting of details. Perhaps more important than any of these is the fact that the program will be used since it is so easy, rather than an engineer's deciding not to make the calculation because it is tedious.

The listing of advantages is misleading unless the difficulties are mentioned. The first few experiences with a computer can be expensive, time-consuming, and traumatic. Computers seem to have personalities of their own. Programs are rarely directly transferable from one machine to another--even of nominally the same type. These are usually transition problems and are worth living through to get the benefits. Engineers in large companies who can turn their problems over to programmers can bypass the time consumption and trauma, but the first bills may induce heart attacks.

The information on the parameter variations of the elements which must be provided to the program can come in many forms: probability density function, minimum and maximum limits, rough engineering guesses of these, etc. The kind of parameter variations analysis that can be done obviously depends on the element input information.

4.4 Amount of Effort

The kind of model, its complexity, ease of analysis, etc., depend heavily on the hardware stage. At first for example, only block diagrams are available for large subsystems. The equations are rudimentary and the analysis is rather primitive.

Not only is the hardware stage important, but the data available and the manpower and facility constraints must be considered. Some thought should be given before starting an analysis to five problems of the resources:

(1) Resources required to plan the program,

(2) Resources required to write down the model or equations,

(3) Resources required to get the data for evaluation,

(4) Resources required for the evaluation itself, and

(5) Resources required for the interpretation of that evaluation.

One cannot say that a certain kind of evaluation must always be done. One can only put a priority on it. Then, according to the resources available, start with the highest priority and work on down the list. If the resources are not enough you can say that certain goals or subgoals cannot be achieved. Then more or less rational choices can be made. With a given amount of resources for the job, obviously planning cannot be allowed to take too much of it.

The complexity of a system sometimes determines how deeply one can go into analyzing it. For example some systems are so complicated that only the most simple analyses can be done, because the very size of the system makes an otherwise simple analysis turn into a very complex one.

All that is required for an anlysis of the variations of the parameters or FOM's of a system is an equation or model of the system in terms of the parameters of the elements and the knowledge of how the parameters of the elements do vary.

5. Techniques for Analyzing Mathematical Equations

Two kinds of notation are used, depending on the circumstances. When it is necessary that one parameter be solved-for explicitly, the equation is expressed* as

$$y = g(\underline{x})$$

and the restrictions on $\underline{x}$ are given for the equation to be true. If the restrictions are not critical or are obvious, they are usually omitted.

Often it is convenient and sometimes it is necessary to use a slightly different functional form. It is more general and does not solve explicitly for any parameter. There is no distinction between dependent and independent parameters, except as one makes it in the course of the analysis.

$$f_1(\underline{x}) = 0, \; f_2(\underline{x}) = 0, \; \ldots$$

In either case, some of the x_i can be random variables, although Sec. 5.1 does not allow this generality. It makes no difference whether the equations are functions of the original variables, or Laplace transforms thereof, or anything else; the mathematical techniques for parameter variations analysis can be the same.

A great many analysis systems are available in the literature, many are given names associated with the purpose of the analysis rather than the technique employed. There are many fewer techniques than there are purposes; so this section is devoted to those few techniques. It is divided into four parts for the analyses:

(1) Probability is not important.

(2) Probability is important; no explicit dependence of probabilities on a common parameter.

(3) Probability is important; probabilities depend explicitly on a common parameter.

(4) Extreme extrapolation of probability is necessary.

and has a final part:

(5) Display of results.

The degree of tractability (ease of manipulation and/or use) of equations varies widely. Tractability is a function not only of the equation itself, but of the manipulator's mathematical ability. There are times when it pays to develop a new function especially if it can be used many times. Developing consists of finding its general behavior and plotting it, finding the restrictions on its argument(s),**

---

* The notation $\underline{x}$ is used to represent a series of parameters $x_1, x_2, \ldots, x_m$. The $\underline{x}$ can be considered as a vector with a finite but unspecified number of components.

** x is defined to be the argument of f(x).

tabulating it, solving for its analytic properties, creating useful computer routines for evaluating it, etc. Once you've done all that, you've turned an intractable problem into a tractable one.

Sometimes the analytic procedures which are indicated in an equation can be easily performed, i.e., the symbols are easily evaluated by simple arithmetic (+, -, ×, ÷) or are available in sufficiently accurate tables or computer routines. Otherwise, when the equation is not tractable, there are several possibilities to try:

(1) Look harder for tables or numerical approximations. This is virtually always the first thing to try, and Refs. 3 and 4 are good starting points.

(2) Linearize the equation--a favorite of engineers for over a century. This can be done by:
   (a) Taylor's series expansion (see Appendix B).
   (b) Mean value theorem (see Appendix C).
   (c) Expansion in series other than Taylor's.*
   (d) Using less complicated models for the items or using ones which are fortuitously compensating.
   (e) Perturbation theory (see, for example, Ref. 5). While in principle this can be used to calculate a second order correction, it is rarely done because of practical difficulties, e.g., the series may rapidly diverge and it may be too complicated and tedious.

(3) Expand the function in a series.* This is the general case for (2) above.

(4) Evaluate the relative magnitudes of the terms and eliminate the ones which are very small. This is a revision of the model and it is usually worthwhile checking to see 'what or how' original assumptions would have to be modified to accomplish the same thing. Further, if some of the terms virtually cancel each other, the process of legitimately neglecting terms can be complicated.**

(5) Where probability density functions(pdf) are concerned, a transformation of some of the variables, either singly or in combination, may allow more tractable pdf's to be used, especially where you are not extrapolating too far out in the tails of the distribution.

---

* Orthogonal polynomials and Fourier series are the most common. Up to the linear term, most all of them give the same result. The advantages of one over the other become apparent when many terms are used.

** Consider as a simple example, $w = x + y - z$ and let $x, z \simeq 1{,}000$, $y \simeq 10$. While y is much smaller than x or z, it is an appreciable fraction of w. Eliminating y would be a catastrophe.

(6) Decide that something else, other than the initially desired operation, will serve your purposes well enough.

(7) Put the problem aside "for awhile". This is probably next most favorite to linearizing.

(8) Call in more expert help.

5.1 If Probability is Not Important

The heading could as well be phrased: "When only probabilities of 0 and 1 are of concern." In this situation correlations (other than 0 or 1) of the parameters are excluded as is the question of statistical independence (see Sec. 2.10).

5.1.1 Direct Calculation

This has always been a popular method of variations analysis where the arithmetic could easily be done. Assume the equations are of the form $f_i(\underline{x}) = 0$. Enough of the parameters are given values (these are called independent-parameters) so that the others (called "dependent") can be calculated. It is assumed that there are several $f_i(\underline{x}) = 0$, so that only a restricted number of x_i can have arbitrarily assigned values. The independent parameters are assigned a new set of values in the vicinity of the first set and the changes in all parameters are calculated. This process can be repeated several times. The use of digital computers makes this method even more attractive.

5.1.2 Series Expansion

When direct calculation is too cumbersome, or when analytic results are desired, or when whim or some other reason dictates, the functions can be expanded into a series. Often the series is infinite and only sometimes can the general term be calculated. Otherwise only a finite number of terms can be known. A Taylor's series is the most popular, but series of orthogonal polynomials have many advantages over the Taylor's series. A Fourier series, also of orthogonal terms, is also quite useful when the function is periodic. A good abbreviated discussion of orthogonal polynomials is found in Ref. 3.

For the Taylor's series, the function is expanded about the nominal point of operations so that $\underline{\Delta x}$ is a direct measure of the parameter variations. Appendix B contains a more complete discussion of Taylor's series.

Most often the expansion is stopped after the first derivative and the equations in Δx_i are linear. It is often advocated, but rarely done, that the higher order terms be evaluated to see if they are really negligible. In any analysis you ought to have a good idea about the error due to dropping higher order terms. It may come from graphs, previous analyses, various estimates of those terms, etc. It can come from evaluating the first derivatives at several points in the region of interest.

If the first derivatives do not vary appreciably, you are safe in neglecting the higher order terms (this is equivalent to numerical differentiation of the first derivatives).

The use of differentials, i.e.,

$$dy = \Sigma_i \frac{\partial g(\underline{x})}{\partial x_i} dx_i \text{ or } \Sigma \frac{\partial f(\underline{x})}{\partial x_i} dx_i = 0 ,$$

is equivalent to neglecting the higher order terms.

5.1.3 Sensitivity

This is a term which is often used in the literature. It has no precise, universally accepted meaning; so don't be dismayed by not knowing it. If you run across several meanings which differ, just remember that it's a free country (more or less) and don't worry about labels. Two concepts which are used quite often are the partial derivative which gives arithmetic variations, and the logarithmic derivative which gives relative* variations.

Consider the arithmetic variations. For simplicity, suppose $y = g(x)$, then the sensitivity is

$$\frac{\Delta y}{\Delta x} \approx \frac{dy}{dx} .$$

In the complex case where $f(\underline{x}) = 0$, the sensistivity of x_i to x_j is

$$\frac{\Delta x_i}{\Delta x_j} \approx \frac{\partial x_i}{\partial x_j} .$$

Now consider relative variations. For simplicity, suppose $y = g(x)$, then the relative sensitivity is

$$\frac{\Delta y}{y} \Big/ \frac{\Delta x}{x} \approx \frac{dy}{y} \Big/ \frac{dx}{x} \approx \frac{dy}{dx} \cdot \frac{x}{y} = \frac{d(\ln y)}{d(\ln x)} .$$

In the complex case where $f(\underline{x}) = 0$, the relative sensitivity of x_i to x_j is

$$\frac{\Delta x_i}{x_i} \Big/ \frac{\Delta x_j}{x_j} \approx \frac{\delta x_i}{x_i} \Big/ \frac{\delta x_j}{x_j} \approx \frac{\partial x_i}{\partial x_j} \cdot \frac{x_j}{x_i} = \frac{\partial \ln x_i}{\partial \ln x_j} .$$

Still another kind of relative variation is possible, e.g., for $y = g(x)$, the sensitivity is

$$\Delta y \Big/ \frac{\Delta x}{x} \approx dy \Big/ \frac{dx}{x} \approx \frac{dy}{dx} \cdot x = \frac{dy}{d(\ln x)} .$$

---

* Also called fractional or percentage.

In the complex case, where $f(\underline{x}) = 0$, the sensitivity of x_i to x_j is

$$\Delta x_i \Big/ \frac{\Delta x_j}{x_j} \approx \delta x_i \Big/ \frac{\delta x_j}{x_j} \approx \frac{\partial x_i}{\partial x_j} \cdot x_j = \frac{\partial\, x_i}{\partial\, \ln x_j} \quad .$$

The nomenclature is rarely the same in several articles, plus it is easy to see that the possibilities are not yet exhausted for kinds of sensitivity. The definitions will often boil down to one of those above, or at least to one which is readily recognizable. As mentioned before, don't worry about consistency in labels nor in trying to match your needs to a label. Just pick something that is suited to your needs, or try several different kinds of sensitivity to see which one is best for the case you have.

The partial derivative notation is ambiguous unless the variables which are independent for the differentiation are shown or easily inferred. Furthermore, if there is more than one $f(\underline{x}) = 0$, the differentiation gets complicated due to there being several dependent variables. See Appendix D, Ref. 6, or any advanced calculus text for a further discussion. Appendix D is rather condensed, but it shows you how to stay out of trouble. Sensitivities are explained further in Sec. 6.1.

Which definition of sensitivity to use is not a matter of right and wrong, but of the degree of utility in the specific application. Use the one that tells you what you want to know. But none of them will tell you everything about the behavior. These sensitivity indices have the inherent disadvantages that they relate only two parameters at once and consider only the linear approximation. If the combined effect of several parameters is desired then some combination must be worked out. But any such combined number, or expression, has the disadvantages that are built into it, viz.,

(1) The particular combination is not unique. There are others that could as well have been used.

(2) One number does not completely describe the multiple effects.

A further discussion of this problem is given in Sec. 6.1.

5.2 Probability Is Important; No Explicit Dependence of Probabilities on a Common Parameter

It is presumed that extreme extrapolation of the pdf (this usually means going far out into the tail region) is not necessary.* Section 5.2.1 treats the case where

---

* This means, at the farthest extrapolation, that cumulative probabilities of less than 1/N, where N ≡ number-of-items-in-sample, will not be considered. It is safer to stay within 2/N or 3/N due to sampling problems. To be quite conservative, you would stay within 5/N or 10/N.

a few moments of the distributions are to be calculated. The rest of the subsections treat the situation where the distribution itself is to be calculated.

There are only a few cases where the problems can be solved analytically, even under simplifying assumptions and even where only limited information is needed. So if you don't hit one of those few, the three choices available are:

(1) Transform the situation to one of the tractable ones. A rather procrustean approach is usually necessary.

(2) Use numerical methods.

(3) Work on something else for awhile, presuming that you have conscientiously exhausted the other possibilities. Any problem which is put off long enough will no longer need to be solved.* If the pressures for a solution get too great, go back to (1) and imitate Procrustes some more.

It is presumed in this section that the population pdf, or enough of its properties, is known. Estimation from a set of data is not treated here.

5.2.1 Evaluation by Moments**

The simplest situation, and the only one amenable to general analytic treatment is where the functions are linear. The equations must be in the form $y = g(\underline{x})$ and, further, must have the special form***

$$y = \sum_i a_i x_i \quad .$$

Introduce the following notation for the special moments--mean and variance (variance is a short name for square-of-standard-deviation):

$\mu \equiv$ average of y, $\quad \mu_i \equiv$ average of x_i;

$\sigma^2 \equiv$ variance of y, $\quad \sigma_i^2 \equiv$ variance of x_i

$\rho_{ij} \equiv \rho_{ji} \equiv$ linear correlation coefficient of x_i and $x_j (i \neq j)$,

$\rho_{ij}\, \sigma_i\, \sigma_j \equiv$ covariance (x_i, x_j).

Then $\mu = \sum_i a_i \mu_i$, $\quad \sigma^2 = \sum_i \sum_{j \neq i} a_i a_j$ covariance $(x_i, x_j) + \sum_i a_i^2 \sigma_i^2$. In words: the mean of a sum is the sum of the means; and the variance of a sum is the sum of the variances plus covariances.

---

* Unfortunately, the phrase "long enough" is defined only by this sentence, thus turning the sentence into a tautology. Unfortunately, it is difficult to give good sound advice without being tautological much of the time.

** Statistical moments are analogous to moments of weight or mass. The r-th moment about the mean (center of gravity or mass) is $M_r \equiv \int (x-\mu)^r p(x)dx$. Pairwise moments about the mean are $M_{rs} = \int (x_1-\mu_1)^r (x_2-\mu_2)^s \, \mathrm{pdf}(x_1,x_2)dx_1dx_2$; M_{11} is the covariance of x_1 and x_2.

*** The summation limits are presumed to be over the appropriate range.

These formulas are true regardless of the pdf's of any of the variables and regardless of any statistical dependences among the variables; in particular the variables need not be Gaussian. There are popular misconceptions about the formulas, so see the following examples; they help to make the formulas clearer.

(1) Let $y = x_1 + x_2 - x_3$

then $\mu = \mu_1 + \mu_2 - \mu_3$

$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2\rho_{12}\,\sigma_1\sigma_2 - 2\rho_{23}\,\sigma_2\sigma_3 - 2\rho_{31}\sigma_3\sigma_1$

(2) Same as (1), but let x_1, x_2, x_3 be linearly uncorrelated*

then $\mu = \mu_1 + \mu_2 - \mu_3$

$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2$

(3) Same as (2), but let x_1 have a Normal pdf, x_2 have the negative exponential pdf, and x_3 a Weibull pdf. Then μ and σ^2 are the same as in (2); these formulas are true regardless of the pdf's.

(4) $y = 2x_1 - x_2 + 3x_3$

$\mu = 2\mu_1 - \mu_2 + 3\mu_3$

$\sigma^2 = 4\sigma_1^2 + \sigma_2^2 + 9\sigma_3^2 - 4\rho_{12}\sigma_1\sigma_2 - 6\rho_{23}\sigma_2\sigma_3 + 12\rho_{31}\sigma_3\sigma_1$

While it is possible to write moment equations higher than the second, it is rarely if ever done in an engineering problem. Some other approach would be used.

If the equation is not linear (and cannot be made that way), then numerical methods must usually be used. Generally, it will be as easy to deal with the pdf's as with a few moments; so the next sections will be applicable.

5.2.2 Evaluation of the pdf's--Tractable Problems

There are not many tractable combinations of probability density functions. The common ones are listed in Appendix F. While it is not necessary that a problem be formulated in a tractable way it is certainly more convenient. Therefore many of the pdf's are chosen so that the result is tractable rather than because one is so much better than the other in describing the situation. Virtually all of the combinations are given for statistically independent variables; statistical independence is sufficient but it may not always be necessary. The completely general joint probability density functions are difficult if not impossible to find for anything but Gaussian distributions.

---

* A sufficient, but not necessary condition is that x_1, x_2, x_3 be pairwise statistically independent; i.e., it will be true if they are pairwise statistically independent and may or may not be true otherwise.

5.2.3 Evaluation of the pdf's--Intractable Problems

The problem is formulated so that

$$y = g(x_1, x_2, \ldots, x_n) ,$$

$$\text{pdf } (x_1, x_2, \ldots, x_n) \text{ is given.}$$

Then

$$\text{pdf}(y) = \int dx_2 \int dx_3 \cdots \int dx_n \, \text{pdf}[x_1(y, x_2, x_3, \ldots, x_n), x_2, \ldots, x_n] \cdot \left|\frac{dx_1}{dy}\right| .$$

where the integrals are taken over the appropriate region of $\underline{x}$. It doesn't matter which variable is taken as x_1. If several variables are transformed, the Jacobian of the transformation must be used; see, for example, Ref. 6.

There are usually one of two difficulties here (if Sec. 5.2.2 doesn't apply):

(1) it is impossible to solve y for any x_i, or

(2) if you can solve it, you cannot integrate it in closed form.

If y is the sum or difference* of several variables, the operation is called convolution. If transforms are used, such as Laplace transforms, convolution of the actual variables is equivalent to multiplication of the transforms. The method of characteristics (in statistics) is equivalent to Laplace transforms.

If the integral is intractable, as assumed in this section, the only way to solve the problem is numerical integration. There are two general kinds of numerical integration:

(1) Direct--the integral is approximated by some kind of sum.

(2) Monte Carlo--instead of a frontal attack on the integral itself, probability is used in the form called Monte Carlo**; one need not invert the function y (i.e., solve it for x_i) to solve the problem, nor must the integral be tractable.

If no multiple integrals are involved, there are a multitude of techniques to use for direct integration. The simplest one generally used is Simpson's rule. The more complex ones are forms of Gaussian integration (Ref. 3, chapter 25 has a short discussion--see Vol. II, Computation, for more complete references to Numerical Methods) wherein the function is evaluated at special, nonuniform intervals.

If multiple integration is involved, the choices are much more restricted. Basically, the multiple integral is replaced by a multiple sum. The multiple sum is then evaluated term by term. If there is an m-tuple integral and n divisions in

---

* Technically its a sum of $x_1 + (-x_2)$.

** The origin of this name is the famous gambling place in the Principality of Morocco; the name was chosen because of the direct association of both activities with chance results.

the range of each variable, there will be on the order of n^m terms in the sum. In principle it is straightforward; in practice it is tedious and time consuming. If this latter is too true, then Monte Carlo procedures should be considered.

AT ANY STAGE, IF THE SITUATION IS TOO COMPLICATED, REFER TO THE 8 SUGGESTIONS IN THE INTRODUCTION TO SECTION 5.

Monte Carlo is an expensive, time-consuming process that is only feasible because of the availability of digital computers. It should be used only when computer time is quite cheap or there is no other way out. Suggestions (6) and (7) in the introduction to Sec. 5 are particularly apt.

If all else has failed and Monte Carlo is the way to go or perhaps you think you should try it at least once to become familiar with the method, then attention must be paid to the pdf of the variables. You have to be able to choose values of x_1 through x_n according to the probability density function. The simplest situation is where the variables are statistically independent and the combined probability density function then directly factors into the product of the individual pdf's.

Computer subroutines are available which effectively divide the interval from 0 to 1 into n equal divisions and make a choice of one of those numbers on an equally likely basis. The individual pdf's are converted into individual cumulative distribution functions (cdf) and the number between 0 and 1 chosen above is converted to a value of x_i by that cdf. A new choice of number is made by the above method for each x_i and an x_i calculated from its own cdf. When the complete set of x_i is available, y is calculated and its value recorded. Very often y is converted from a continuous variable to a discrete variable where, for example, the probability of success or failure is of importance.

The process is repeated tens, hundreds, or thousands of times depending on the accuracy desired, the machine time available, and the money allowable. Roughly speaking you will not be able to estimate probabilities smaller than 1/N where N is the number of Monte Carlo trials.

If the variables are not statistically independent and a linear correlation coefficient will describe the statistical dependence adequately, techniques are available in the literature for choosing the probabilities taking these linear correlations into account.

If a more complicated statistical dependence than simple linear correlation is necessary, there will be severe practical difficulties in calculating the x_i and professional help will be needed. As before, the advice to simplify the problem should be taken to heart.

The combined distribution function of the variables regardless of how it is made up is rarely known very accurately, i.e., the correspondence of a given equation

to physical reality is usually rather dubious especially out on the tails; therefore, there is a practical limit as to the number of trials to make because the increased precision will merely result in more specific knowledge about your ignorance.

There are several ways of analyzing the resulting data. One way, for example, is to draw a cumulative distribution curve for the y and use it to calculate the tolerance intervals. Another way is to calculate the first several moments of the distribution of y and to find one of the tractable distributions which fit the data reasonably well. In this procedure care must be used to distinguish between statistical and engineering significance. For example, even though the data may obviously not fit a Normal distribution from a statistical viewpoint, they may well fit one close enough for the engineering purposes at hand. It is very likely that professional help from a statistician will be necessary during this process but the engineer should never abdicate his responsibilities in this area anymore than any other area. He should learn the whys and wherefores as much as necessary and then make his own decision. A little practice with asking questions will yield great benefits. Do not be bashful--you are not getting paid to have a complete understanding of the statistical literature.

If there is more than one parameter y to be calculated from the x_i, the same set of x_i may be used for each one of the y's--if the fact that the results on the y's will be highly correlated will not be of any importance. This will have the advantage of saving some machine time.

Special techniques for calculating an x_i from the probability between 0 and 1 are available for some distributions and a specialist in numerical analysis can be of assistance here.

It is not expected from the descriptions given here that one will go out to a computer and perform the calculations himself. This discussion is intended to make the engineer able to interpret his problems more accurately and confidently to the numerical analyst or computer programmer.*

5.3 Probability Is Important; Probabilities Depend Explicitly on a Common Parameter

The explicit dependence of the probabilities of several variables on a common parameter can take several forms:

(1) The type of distribution stays the same as the time changes. (This common parameter is referred to as "time" throughout this section, although it can be

---

* There is some specialization in this field. A Computer Programmer will take the particular operation you wish to perform and put it on the computer to be used. The Numerical Analyst will decide from the general analytic description of the problem what kind of numerical approximation to make. Very often these specialties are not combined in the same person.

anything else.) The parameters of that distribution are then functions of time, and the values of the random variable at different times are statistically independent. In a few cases it may be possible to get an analytic solution of the problem as a function of this common parameter but these cases would be extremely rare and the occurrence would probably be most fortuitous.

(2) The general situation is as in (1) above, but the values of the random variable are statistically dependent. Such might well be the case, for example, for distributions describing the drift of resistors. It would be most unlikely to have the value of a resistor at one time statistically independent of its value at another. If this statistical dependence exists, the analysis is likely to be very complex; the services of a statistician and/or mathematician will probably be necessary. Just keeping track of everything exactly, and meaning exactly what is written down may be difficulties enough for one person.

(3) If the probability distributions are discrete the extra parameter can perhaps be treated as another event.

The techniques of analysis will be the same as in Sec. 5.2 but more complicated and complex. Further simplifications will probably be made in order to handle the problem. The thing to watch out for here is that the simplifications are not too simple-minded.

5.4 Extreme Extrapolation of Probability Is Necessary

If you need to read this chapter you are already in trouble--extrapolating out into the tails of probability distributions is one of the most hazardous kinds of extrapolation there is. The reason for including this section in the volume on parameter variations analysis is that one of the regions of most concern in high reliability is where the parameter has a value way out on a tail of the pdf, viz., has very low probability of occuring.

5.4.1 A Criterion for Rejecting Out-liers Is in Use

An out-lier is a data point that lies way out from the rest of the data. Some people believe such points can be rejected on purely statistical grounds, others do not. The case where a datum is rejected, regardless of its location, on sound physical grounds is entirely different and not considered here. There are good reasons for being suspicious of any data rejection which is made on statistical grounds, but this is of concern in Sec. 5.4.2, not here. In that section, a good explanation is given of sorting the data into two categories, but still using all of it.

Assuming that, in fact, out-liers are being statistically rejected, a useful rule of thumb for keeping the situation under control is:

Never extrapolate any further than you are willing to accept a datum without calling it an out-lier.

An example of this is shown in Fig. 1 for Chauvenet's criterion.* It is used in the example because it is fairly common and is one of the better criteria for rejection of out-liers. The criterion has been slightly modified so that it more easily fits the scale of plotting positions (N+1 has been substituted for N); the plotting positions are expected values.** With expected value plotting positions, the first point is plotted at 1/N+1 (N is the number in the sample), the second point at 2/N+1, etc. Chauvenet's criterion for rejection of an out-lier can be stated as:

(1) Draw the estimated cumulative probability line through the data points on cumulative probability paper.

(2) Find the probability corresponding to ½/N+1.

(3) Find the intersection of that probability line with the calculated line.

(4) This value of ordinate is the dividing line between rejectable out-liers and good points.

In the series of points marked with X on the graph there are no out-liers; in the series of points marked with □ on the graph the first two are out-liers.*** The estimated line is shown dashed beyond the out-lier limits and solid lines are extended parallel to the probability axis. Since you have now rejected any and all information from points in the "strength" region outside the out-lier criteria you have no information about the distribution in that region and it is not possible to extrapolate the distribution validly in that region. Instead in the example in Fig. 1 the cumulative distribution curve is drawn so that the probability is a constant anywhere in the out-lier region. Methods associated with the Chebyshev**** criterion for showing

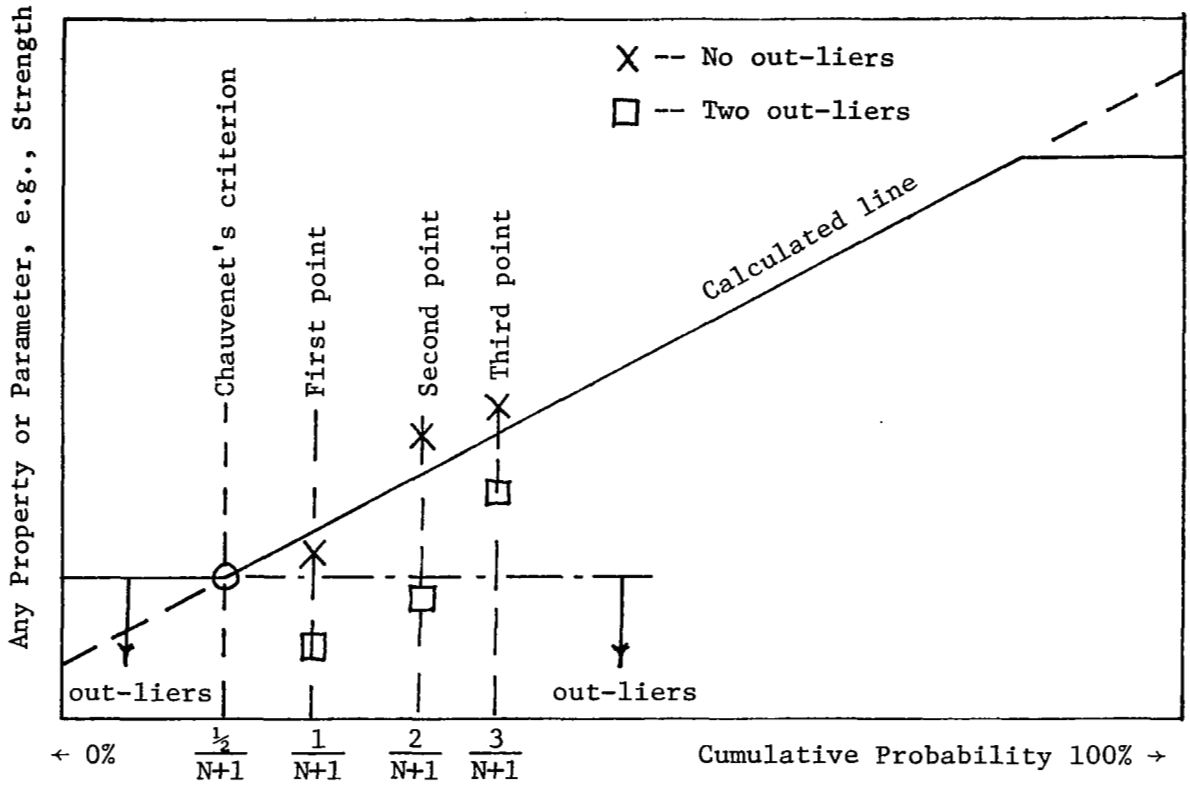
---

* A Normal distribution with N=9 is used to plot the illustration.

** There are other possible plotting positions. Many authors have their favorites. If it turns out to make an important difference to your results which one you use, you are in serious trouble because the uncertainty due to scatter is much more than that due to plotting position. If it makes negligible difference which one you use, which is the case when there are a large number of points, then of course there is no trouble.

*** We assume for illustration that either set of points would give the same calculated line.

**** This name has many spellings in English since the transliteration from the Russian is not unique. Any phonetic spelling is as "proper" as any other.



Extrapolation When There Is an Out-lier Criterion

Figure 1

limits on the ignorance region are discussed in Sec. 5.5.2. Suffice it to say here that they are usually unsatisfactory because they are too weak.

For convenience call the region which contains the out-liers the defective region. Then the distribution can be considered as made up of a tractable one in the central region plus an unknown one outside that region. The central region is usually reasonably well taken care of by the data and the problem becomes one of trying to estimate the fraction in the unknown region which, for simplicity, will be called the fraction defective (fd). The true fd can be estimated by usual quality control techniques, e.g., a single sample.

Very often the sample size will be small enough so that the estimated fd is dishearteningly large for any reasonable confidence even for no defectives found in the region. In this case the engineer should consider the uses of prior information, although in critical applications one is hard put to it to find a satisfactory method of using any kind of prior information. In other applications an engineer will generally use his own judgment and he will have a tendency to over-estimate the goodness of the situation.

The two very basic difficulties of an out-lier criterion for high reliability applications are:

(1) If data are outright rejected in some out-lier region then you can't tell at all what's going on in that out-lier region. But that region is precisely the place where you must be making estimates about the situation.

(2) If data are rejected in the out-lier region only for purposes of estimating the central distribution, and then if the number of outliers are used to estimate the fraction lying in the out-lier region, the resulting estimate of fraction defective is too large to be of use.

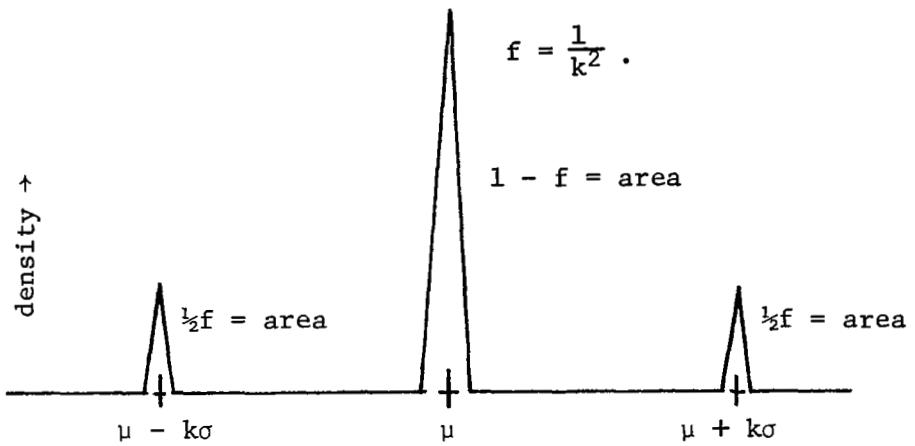
5.4.2 No Statistical Rejection of Data Points Is Used

When there is no criterion for rejection of data points, you have at least a fighting chance for estimating low probabilities since they are not excluded by the nature of your criterion. In this kind of situation one sometimes resorts to the Chebyshev approximation which states that IF the true mean (μ) and true standard deviation (σ) of a sample are known, then the fraction which lies in the tail region beyond $\pm k\sigma$ is less than or equal to $1/k^2$. Most often this is a dishearteningly large fraction and it pays to see in a geometric way how it arises. Variance corresponds to the physical moment of inertia and, for a given mass, the maximum moment of inertia is obtained by having all the material concentrated at the outer boundaries. In the Chebyshev case the distribution will be two spikes each of area $\frac{1}{2}f$ at $\pm k\sigma$ from the mean and a spike of area 1-f at the mean as shown in Fig. 2. The calculated

Variance Equation:

$$2 \times \tfrac{1}{2}f \times \overline{k\sigma}^2 + (1 - f) \times 0 = \sigma^2 ,$$

$$f = \frac{1}{k^2} .$$



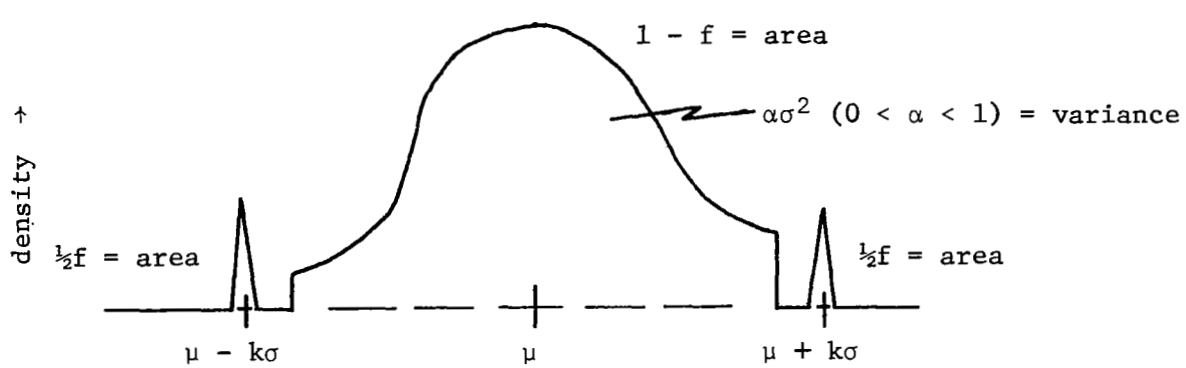
Worst Distribution of Mass

Chebyshev Criterion

Figure 2

Variance Equation:

$$2 \times \tfrac{1}{2}f \times \overline{k\sigma}^2 + \alpha\sigma^2 = \sigma^2 , \quad f = (1 - \alpha)/k^2.$$



Worst Distribution of Mass

Modified Chebyshev Criterion

Figure 3

variance is $2 \times \frac{1}{2}f \times \overline{k\sigma}^2 + 0 = \sigma^2$. Therefore the amount at a distance $\pm k\sigma$ from μ is

$$f = 1/k^2$$

for this worst possible case.

Often one has much more information than is used in the previous derivation. If so, he can calculate a modified Chebyshev criterion.* Assume that within certain limits of the variable, its distribution can be represented by some tractable one. What, then, is the worst possible situation in the tails? It is shown in Fig. 3 that: IF the true mean and true variance are known and if some of the distribution is known to lie in a central region with the mean of the central region at the true mean, then the amount at a distance $\pm k\sigma$ from μ is

$$f' = (1 - \alpha)/k^2$$

where α is the fraction of the total variance that is included in the central part of the distribution. If α were 90% to 99%, for example, f is much less than given above, and is less likely to cause frustration for the engineer. The unfortunate part of the Chebyshev and modified Chebyshev inequalities is that they assume that the true mean and the true variance are known. The importance of this assumption cannot be overemphasized, especially in the modified Chebyshev criterion. Unfortunately, all that is ever known (and it is only estimated) is the variance and mean of the central portion; it is precisely those pips out on the tails--so unknown and elusive--that modify the variance ever so slightly but ever so importantly. Therefore the two Chebyshev formulas are useless for estimating the area in the extreme tail region. Since this is not the usual position taken in the literature each engineer should convince himself of the applicability of the above reasoning (i.e., the previous two sentences).**

It is shown in Appendix A that the change in the mean due to a pip on the tail requires many more data to detect*** than does a change in the variance. It is also shown that the number required to detect*** a change in the variance is $2/k^4f^2$. This

---

* This procedure is not common in the literature and Chebyshev had nothing to do with it (as far as is known)--but it is similar to the original Chebyshev derivation and reduces to it for nothing known about the distribution.

** The reason the discussion is included is to show what is bad as well as what is good.

*** If the change is to be detected the uncertainty in the estimate due to finite sample size cannot be larger than the change due to the pip.

is graphed in Fig. 4. As an example, 10^{-4} defectives at 10σ would require over 20,000 observations in order to detect the difference in the variance due to the pip. Of course, if an engineer were to find one of those observations at 10σ he would undoubtedly discard it for all sorts of good reasons. This paragraph confirms the statement made in Sec. 5.5.1 that we cannot know the variance accurately enough to use the modified-Chebyshev inequality.

Eventually, after searching around for a method to use for extrapolation, you will be forced to the one discusssd above in Sec. 5.5.1 except that the "bad" data are not called out-liers but are used to estimate the boundaries of the central region and to estimate the fraction in the "defective" regions.

5.4.3 Extreme Extrapolation Is Necessary

When extreme extrapolation appears to be necessary, measures should be instituted to avoid it. Suggestions are:

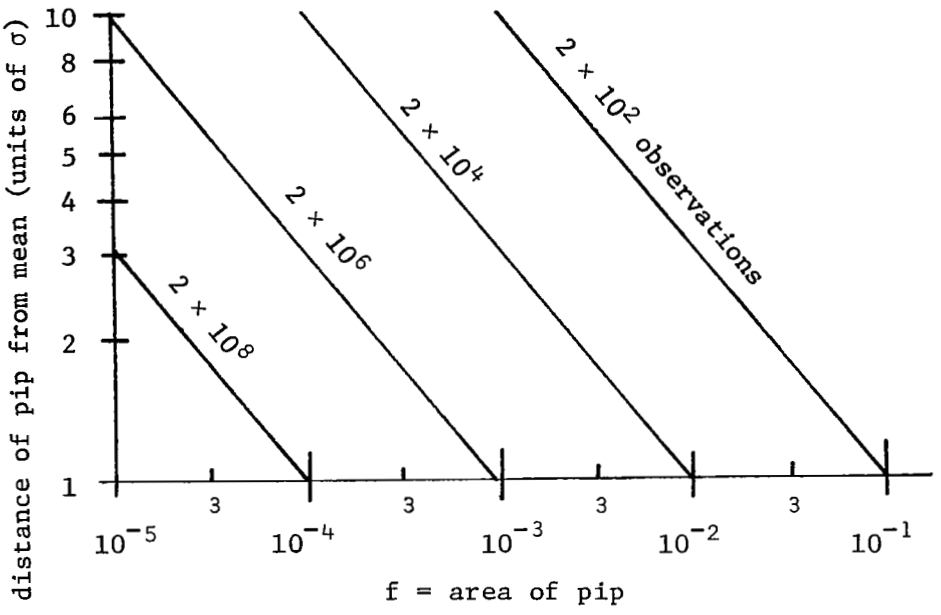
(1) Use a safety margin.* Estimated values of 3 to 6 are often used for the safety margin. Anything much less than 3 is not likely to be enough unless fraction failures on the order of 1% or more are satisfactory. Safety margins of more than 6 are likely to result in overdesign because the accuracy of the model is usually not that good; that is, in order to get that much safety margin you may well have to change some of the properties of the materials, use a different kind of material, or change the design. The failure modes you have neglected, or the things you are not accounting for in the new setup, or the approximations that you are making in the calculations may well negate the extra safety margin benefits. For example, if steel is made harder to make it stronger it may be more susceptable to brittle fracture, its notch sensitivity may be drastically increased, its fatigue strength may have been lowered, its corrosion susceptibility may be much higher, etc. It is best not to calculate a probability associated with these safety margins since the distributions are usually not known well enough.

(2) Use a 100% screen on the material to eliminate those items which otherwise would be in the tail region.

(3) Try to eliminate some of the uncertainty by using ordinary engineering formulas to calculate the deviations. (The formulas are ordinary but their application to this is not.)

(4) Test a great many more samples. This is even more difficult than it sounds since if you are to make an estimate about the following year's production, you have to have a random sample from that production.

* Safety margin is defined here as the difference in true means divided by the standard deviation of that difference.



Number of Observations to Detect the Change in Variance Due to a Pip on the Tail of a Distribution (See Appendix A)

Figure 4

(5) Use nondestructive testing in combination with the engineering analysis of (3) to eliminate the poor elements.

(6) If there is a contractual or other firm requirement that a probability must be associated with a certain extrapolation, practically anything (and anything practical) you do will be a lie. But using the Gaussian distribution will probably get you by with no serious questions and will satisfy the requirement--just avoid kidding yourself even though you are kidding someone else.

(7) Try to make at least qualitative use of your prior knowledge of the elements and how they are made. Even though it is difficult to take into account statistically and quantitatively, it is still useful information for engineering purposes.

Remember that just because a distribution is tractable does not mean that it is practical. When someone says, for example, that an ideal-batch has a Gaussian distribution remember that he is defining ideal-batch rather than giving a description of a previously defined ideal-batch.

In some cases data are recorded on the near tails. For example, 1,000 items may be put on life test and the first 10 failures recorded. In this case extrapolation is not as far into the tail region as if the first 10 out of 100 were measured. If all of the sample has been failed or measured it is wise to remember that only the weaker ones are going to affect the reliability. Therefore, extrapolation is best made using only the bottom half or less of the points.* Some techniques are available for fitting a line to these censored or truncated distributions and these have to be used unless an eyeball line is drawn through the points.

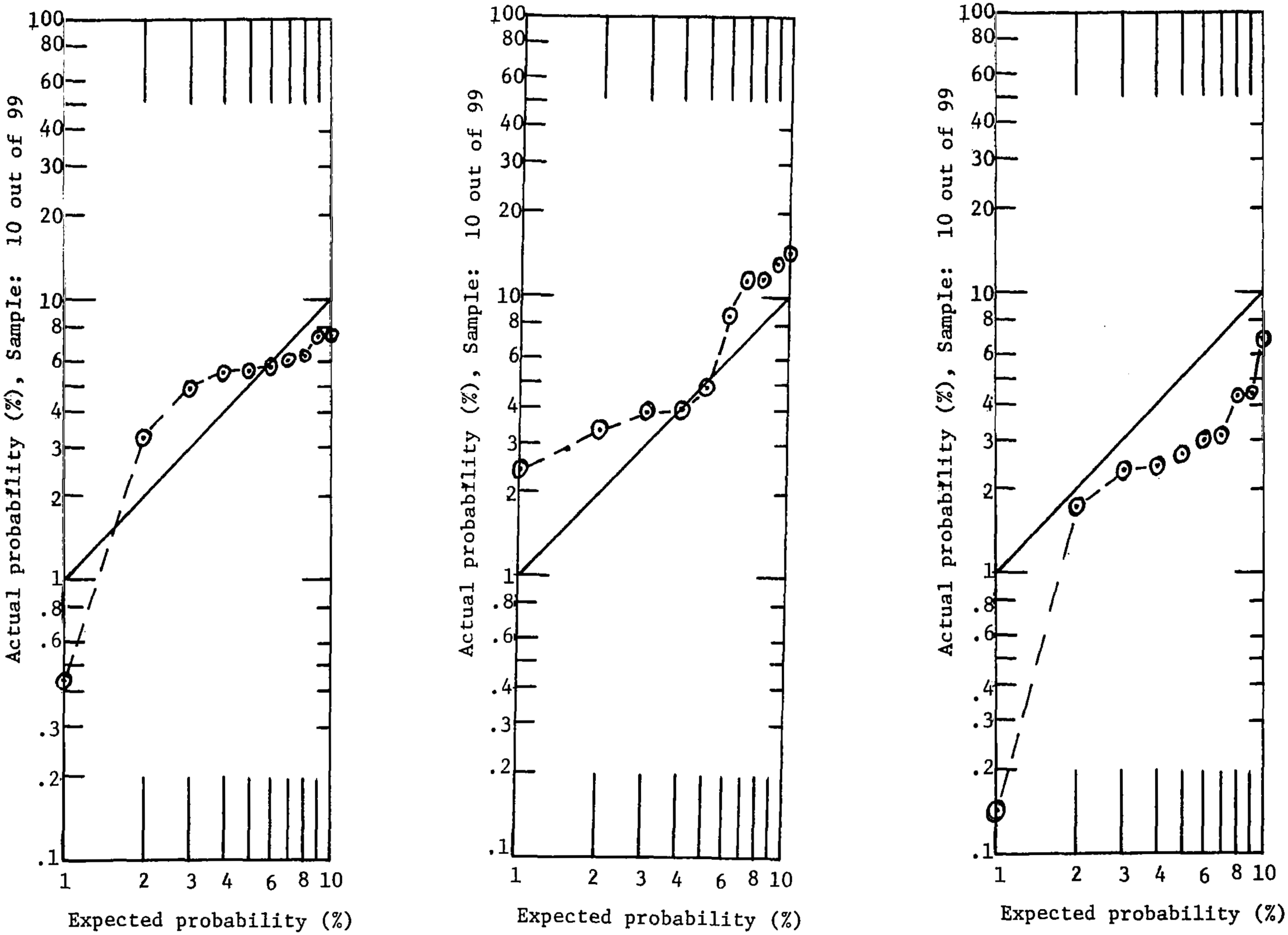
5.4.4 Small Random Samples Are Unreliable

In Figs. 5 and 6 the results from random samples are shown. In all cases the numbers are probabilities directly without having been fitted to a typical distribution, i.e., random numbers are picked in the interval from 0 to 1 with a uniform probability density.** The method of presentation shown here is relatively distribution free. The "low probability" graphs (lowest 10 out of 99 points--Fig. 5) are plotted on log-log paper to enable the large range of numbers to be shown; the "high probability" graphs (all 9 out of 9 points--Fig. 6) are plotted on Gaussian-Gaussian

---

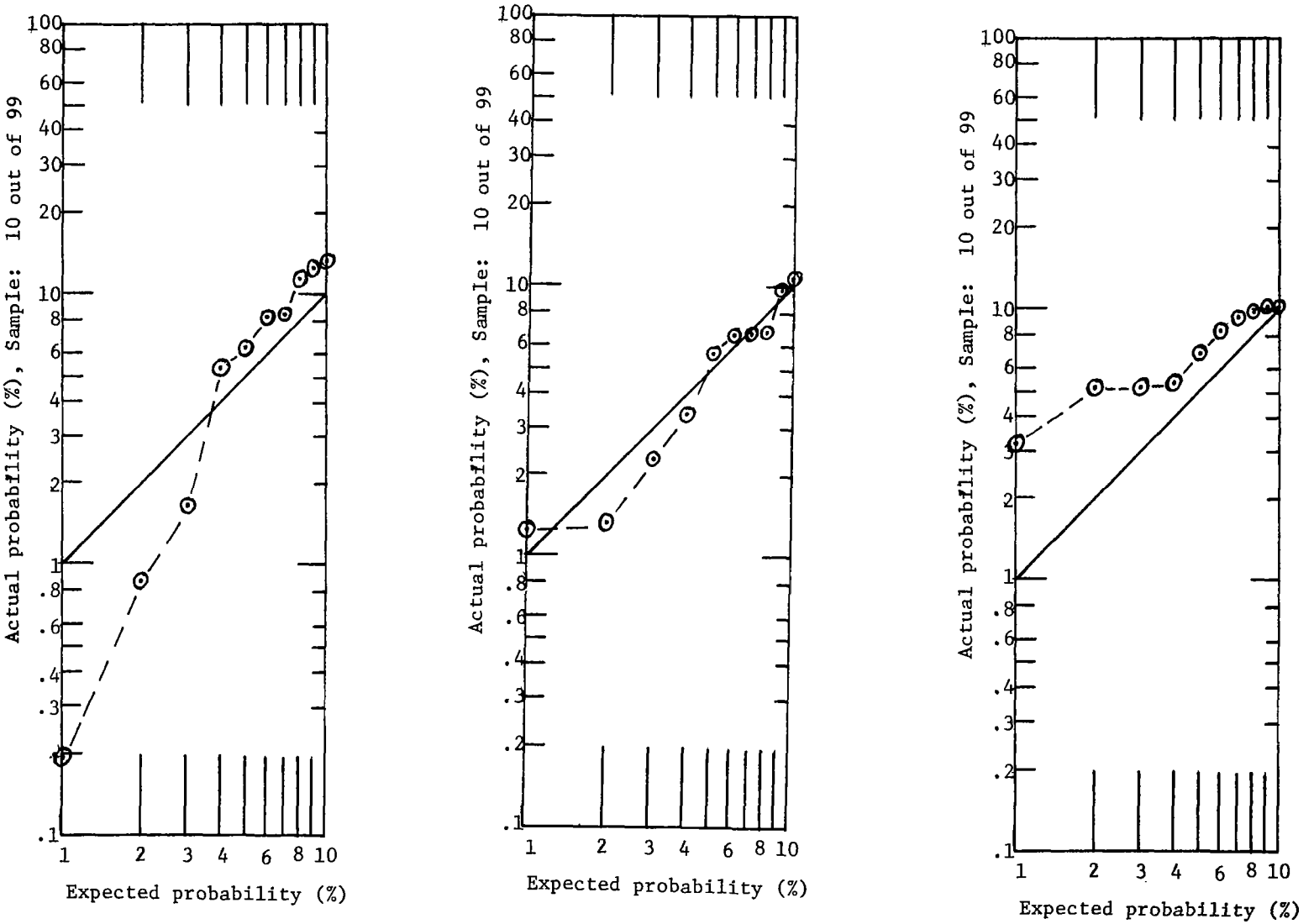
* The upper half of the points can be analyzed to see what made them good and used as engineering feedback to design.

** These numbers are all generated on a computer by one of the algorithms for generating pseudo-random numbers. The exact technique is not important here. (It can easily be argued that no machine program method can generate truly random numbers.)



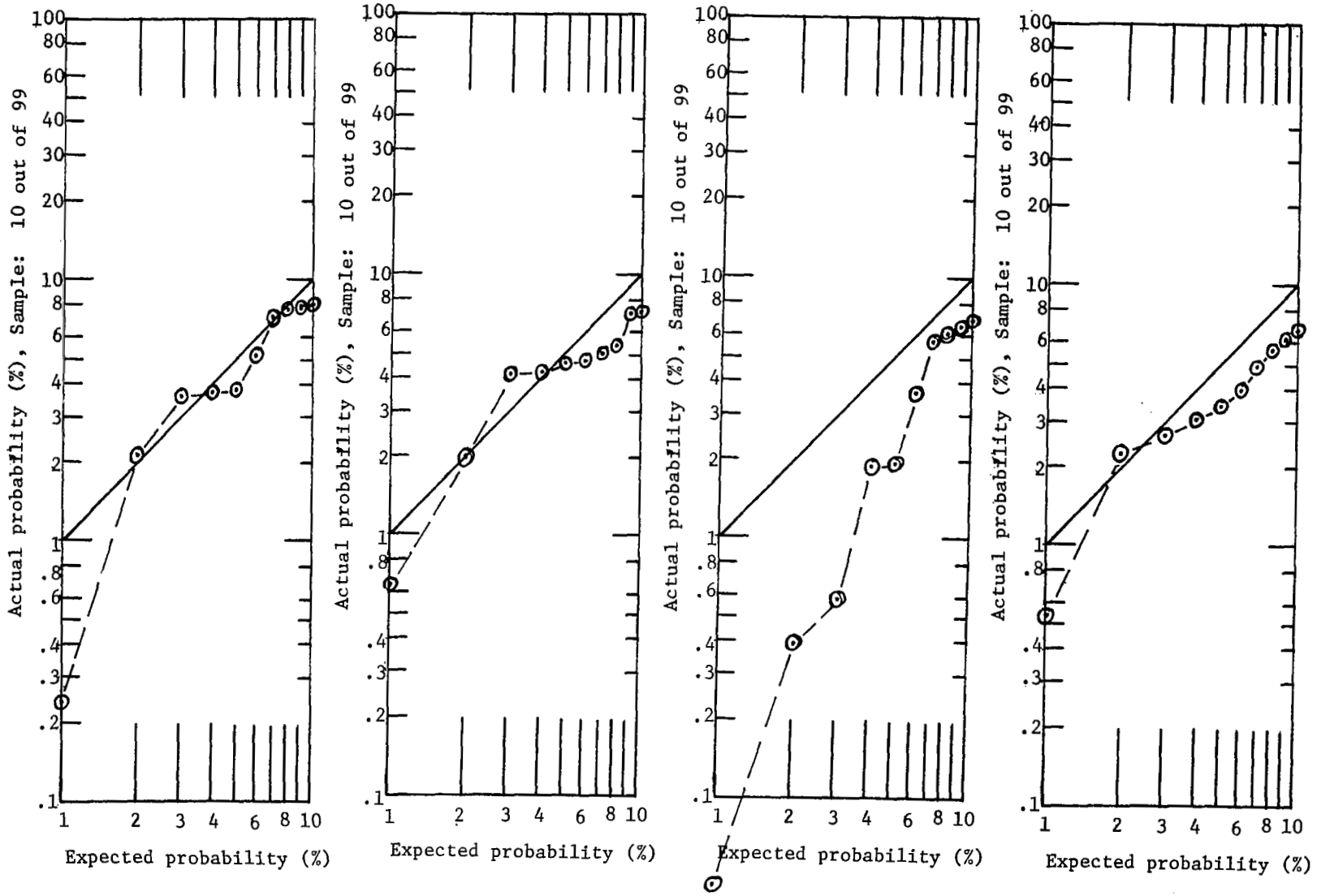
Random Samples

Figure 5a



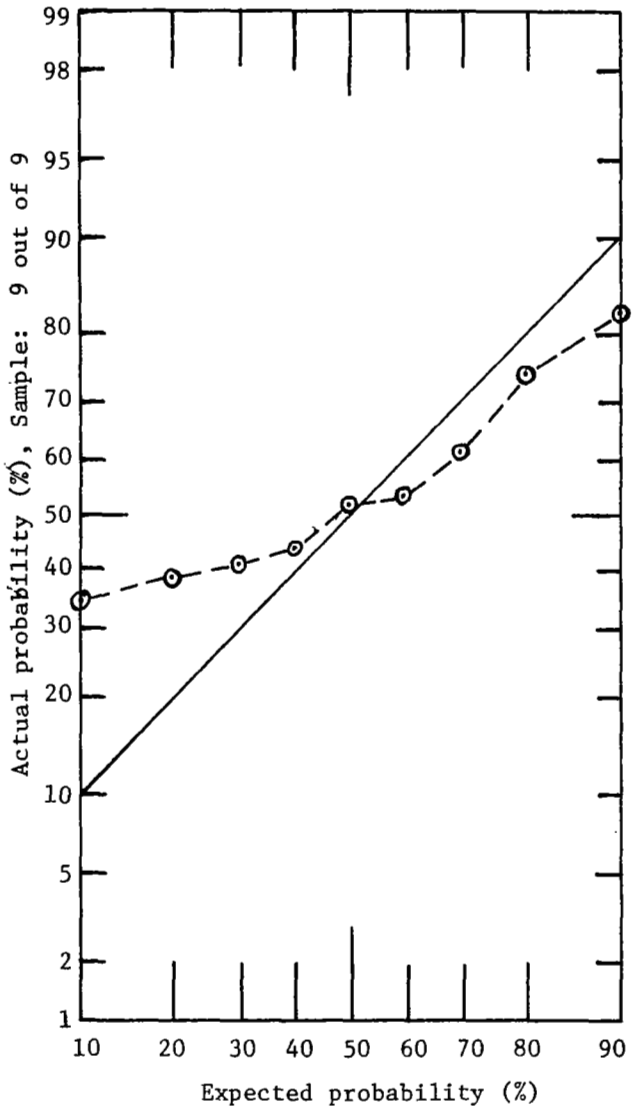
Random Samples

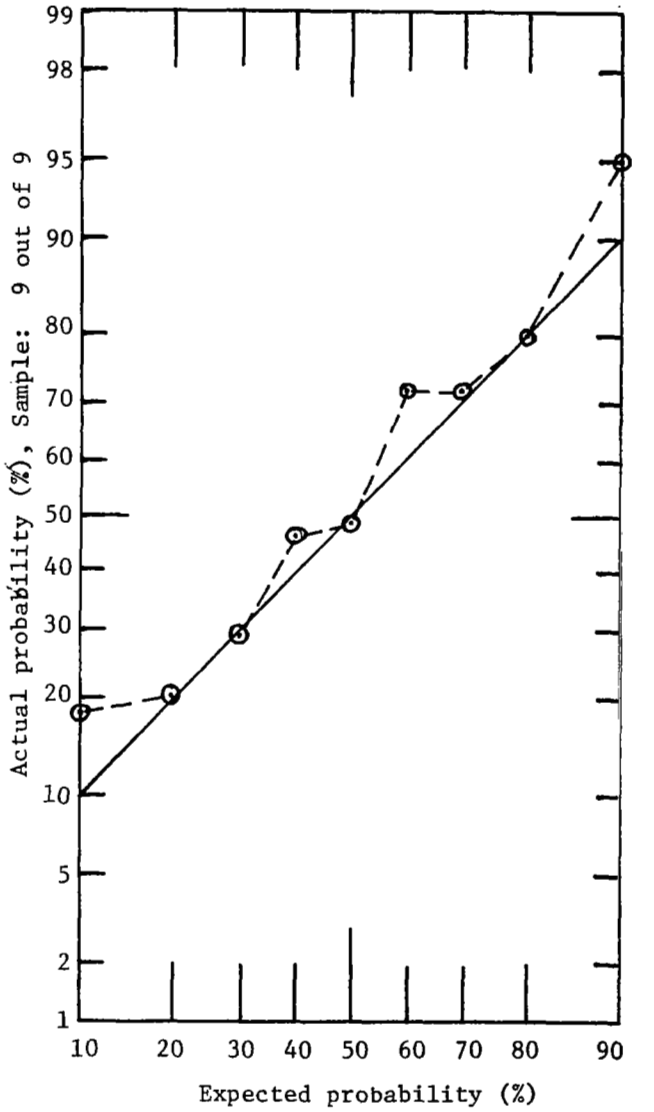
Figure 5b



Random Samples

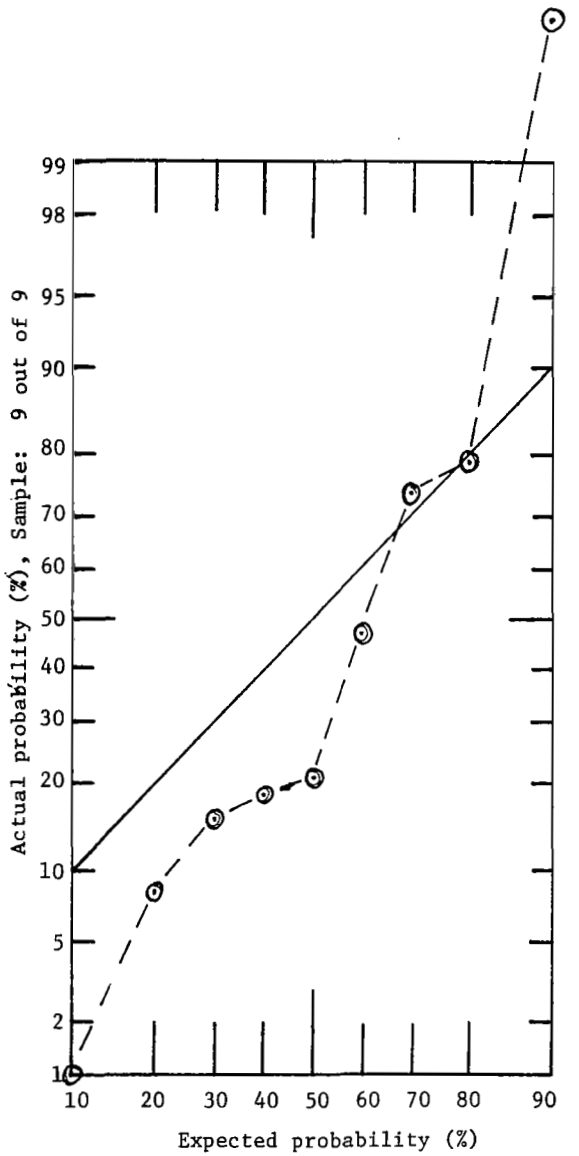
Figure 5c

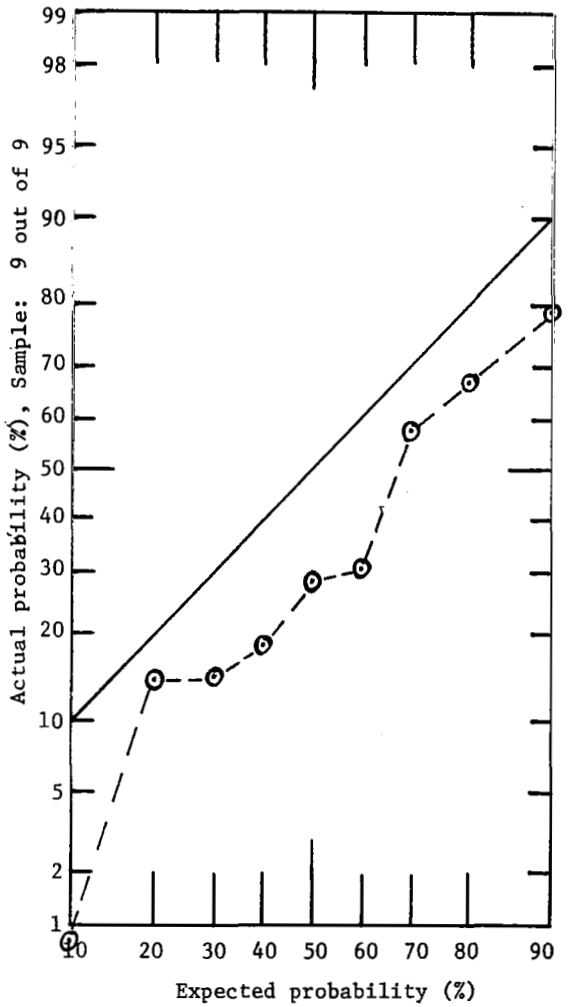




Random Samples

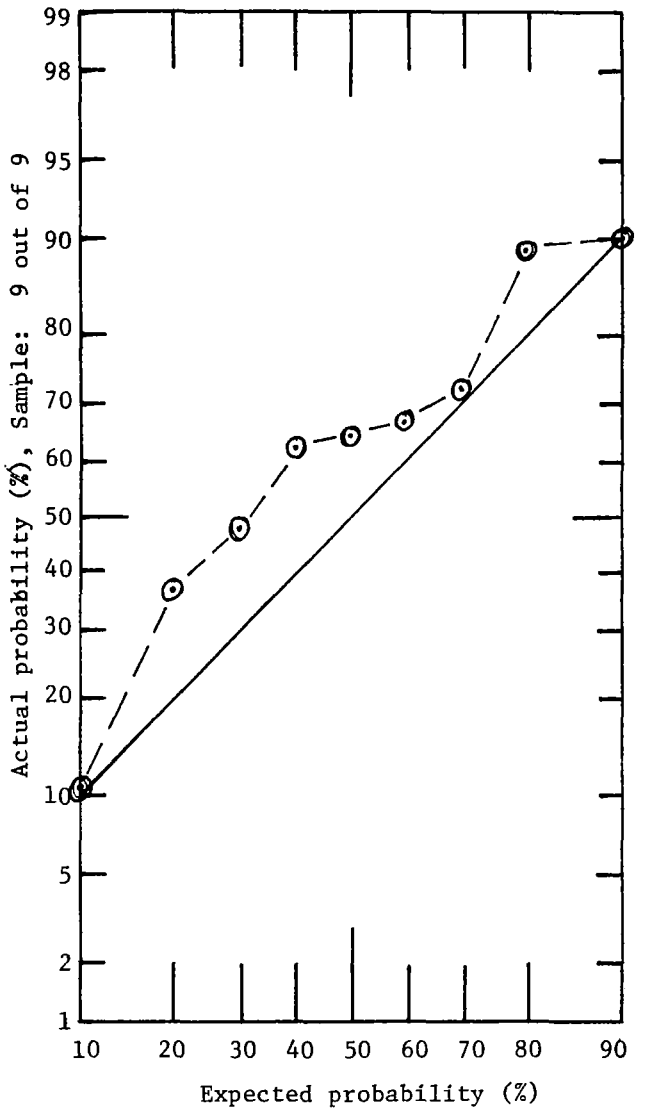
Figure 6a

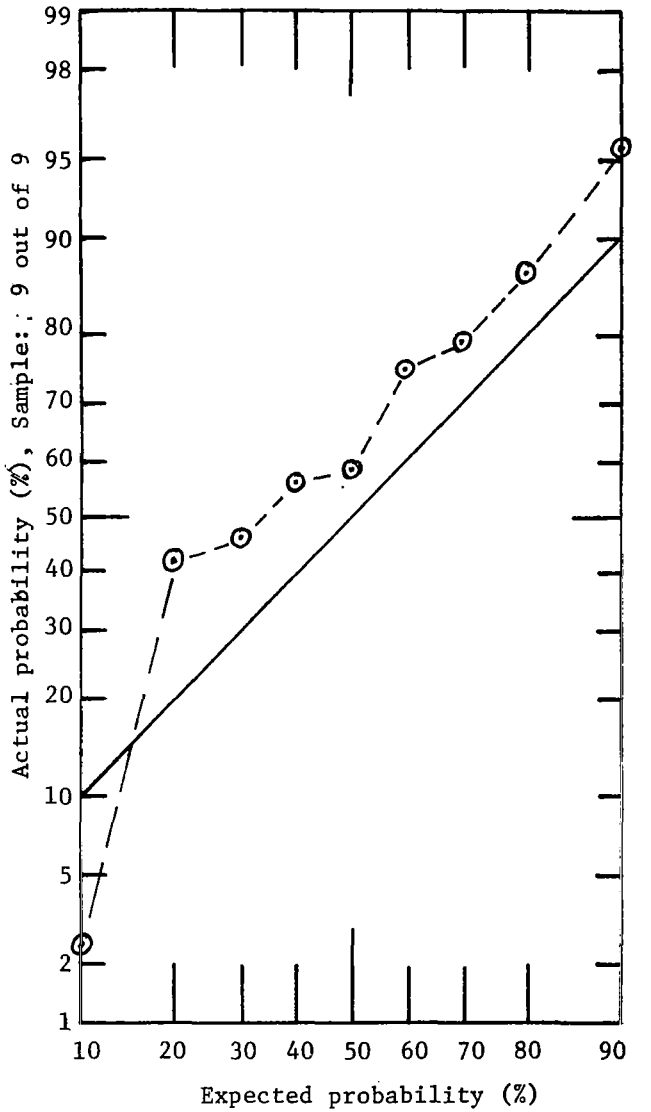




Random Samples

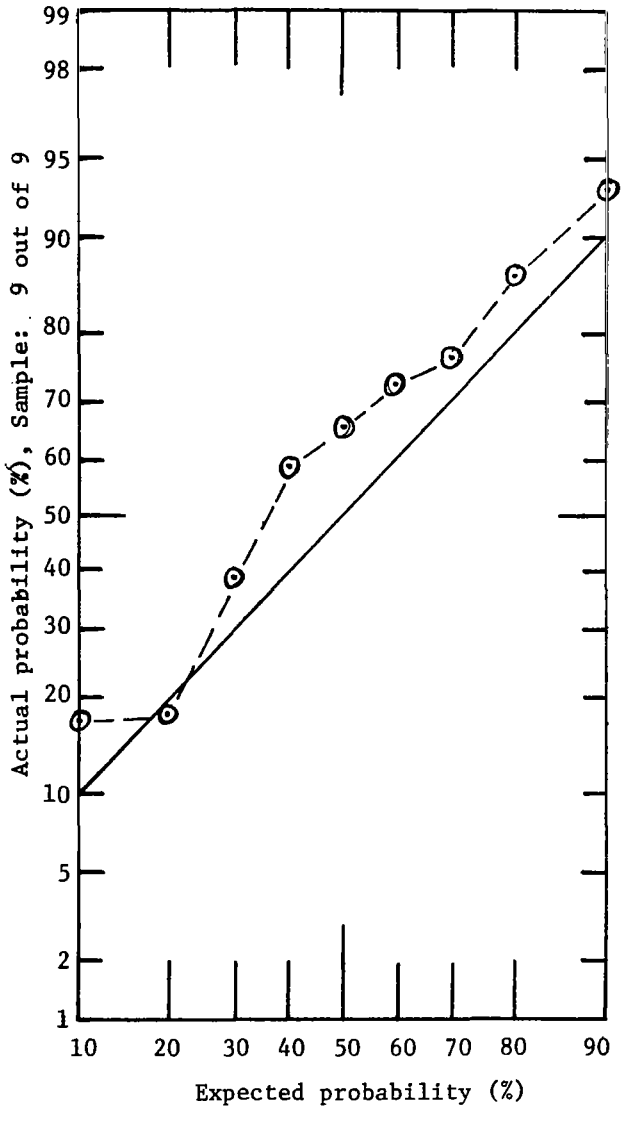
Figure 6b

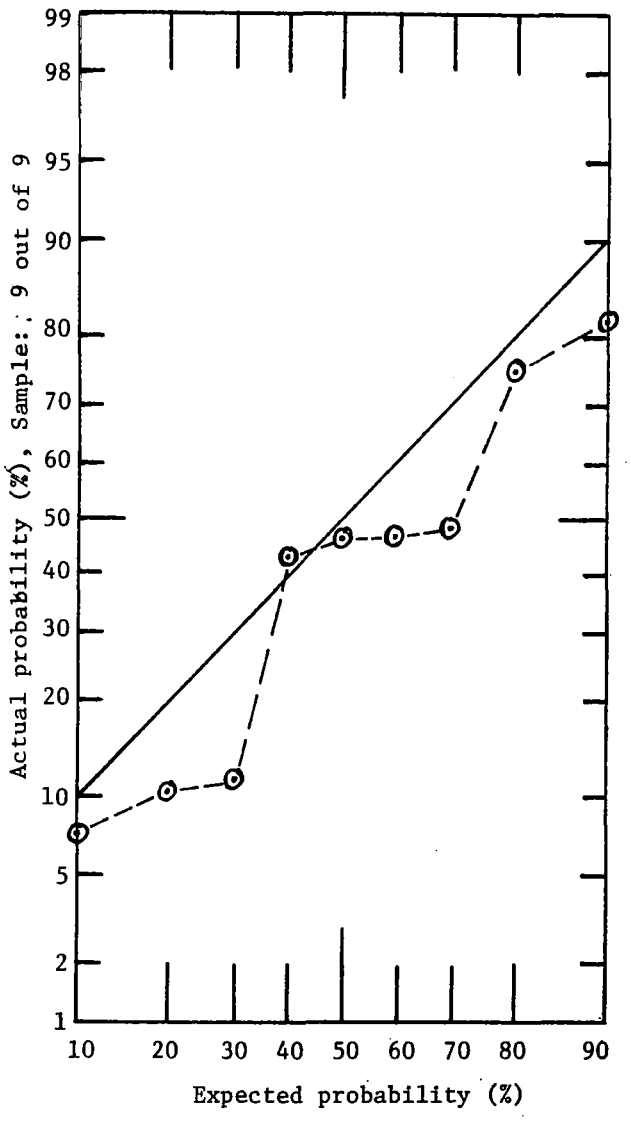




Random Samples

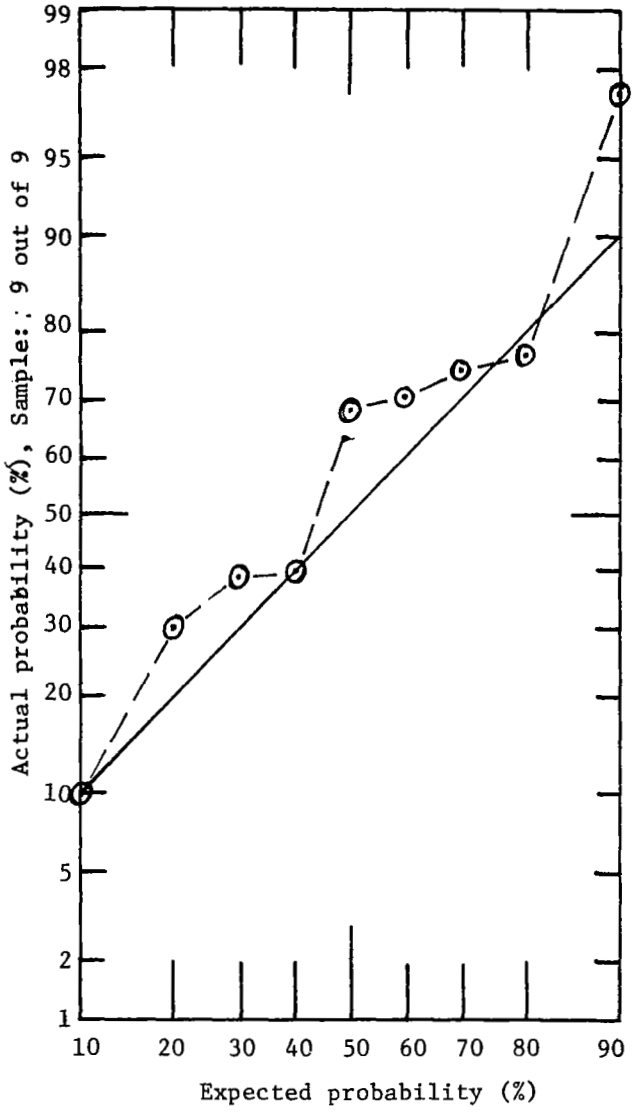
Figure 6c

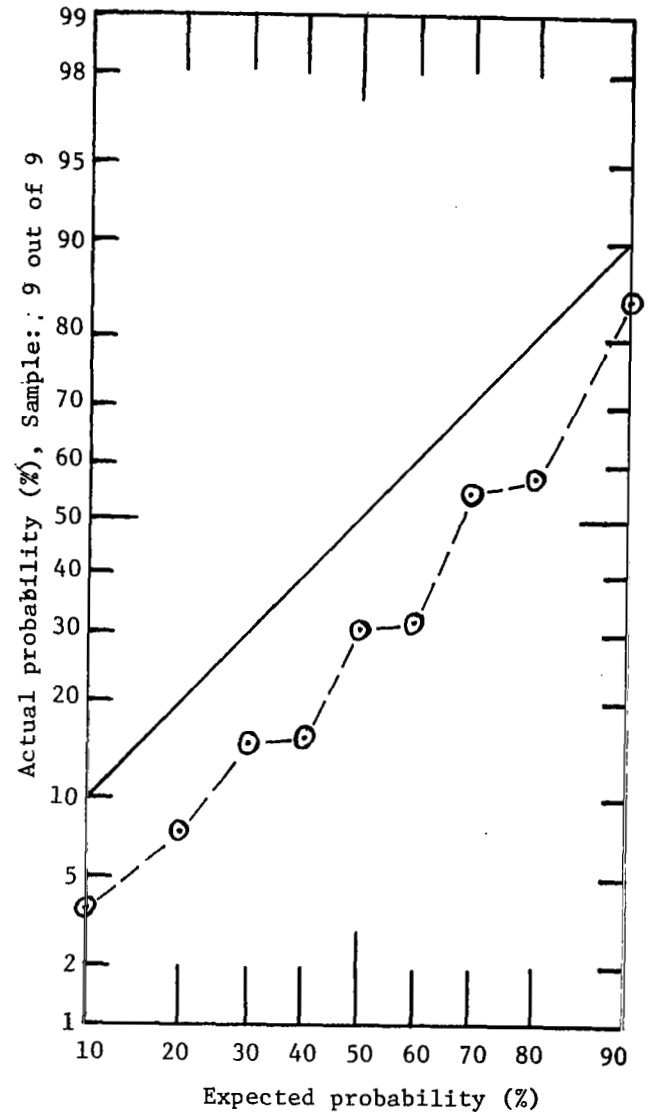




Random Samples

Figure 6d





Random Samples

Figure 6e

scales to better show the entire range. (The exact appearance of the graph depends on the scale used to represent the numbers.) In the limit, all of the numbers would lie on a 45° line. The actual randomly-selected probabilities are plotted versus the expected values (i/N+1 where i is the i-th ordered point from the bottom and N is the sample size). In the Fig. 5,10 samples of the 10 lowest out of about 100 points (99 to be exact; so the plotting positions are round numbers) are plotted to show the extreme variation to be expected in the tail region and to show why guessing at the distribution from such a sample can be so misleading. In Fig. 6, the sample size is 9 and all 9 points are shown. The fact that the most of the point sets do not look as if they were supposed to fall on a 45° line through the origin is significant because a very large sample would look that way. Again note how misleading it would be to estimate the true line from most of the samples.

The problem of out-liers could be treated on these figures. A line is drawn at the "expected plotting position, ½/N+1" to correspond to a Chauvenet criterion. The points which lie below it would then be considered out-liers or defectives.

The purpose of the figures is to illustrate what random samples can look like. The line at 45° is the true population line, why not sketch in your own eyeball estimated line? With variations like those shown it is easy to see how one can be misled by a sample and why extrapolation into the far tails is poor practice.

5.5 Methods of Display of Results

There are four basic methods of display of the results:

(1) graphs,

(2) equations,

(3) tables,

(4) computer routines.

The one to be used is governed by the needs of the occasion and by the time, money, and equipment available--and by personal whim.

There are several forms of graphs. In the usual kind a dependent variable is plotted versus an independent one with, occasionally, a second independent variable being used to label some separate curves. Examples of this are collector voltage-current curves for a transistor with base current as the second independent variable, and pressure-flow curves for a centrifugal pump with motor speed the second independent variable. Even less occasionally a third and further independent variables can be used on separate curves. Very occasionally someone makes a three-dimensional graph. Another kind of graph shows contours: the x and y axes are used for the

independent variables, and contours for given values of the dependent variable are shown. Contour graphs are sometimes called Schmoo plots in electronic engineering circles.*

Equations are a very common form of displaying the results of an analysis, especially if they can be put in a compact tractable form. If the equations are infinite series or very complex it is usually helpful in addition to use one of the other display methods.

Tables are used in preference to or to supplement graphs where the additional accuracy is needed or where many values will have to be looked up.

Computer routines are a relatively new form of display of results wherein a computer program for calculating the results is given but the results per se are not. This is very helpful of course, much more so than tables, graphs, or written equations, when the results must be used within other computer programs. For example, a computer routine for calculating a particular function to four significant figures is much more useful in a computer analysis than is a table of values. If extensive use is to be made of the computer routine, it is often advisable to prepare one of the empirical interpolation formulas such as a Fourier series or a Chebyshev series in order to be able to calculate more efficiently the values to a given accuracy. In this case the original form of the equation is lost but the numerical accuracy is preserved.

Fairly often it will be desirable to use more than one form of output, perhaps even all four in order to have the information available in the easiest way to different people who wish to use it. Sometimes there will be money and calendar-time problems if the display of results is too fancy.

---

* The reason for this is presumably the resemblance of many contours to the Schmoos of Lil' Abner fame.

6. Uses of Mathematical Analyses

There are only a very few basic techniques for analyzing mathematical models but the models are used in a great variety of ways. Much of the literature is oriented toward a discussion of the purpose of the model and the names of a particular analysis are often associated with that purpose, e.g., worst-case analysis. This section will be short since the basic techniques are given in Sec. 5. When you run across a "new" analysis technique in the literature, ask the question: Is it named for the purpose or the mathematical method? There are many purposes and few methods. You are likely to find that the method is a standard one.

6.1 Sensitivity and Tolerance Analysis

These are not precisely defined terms, but they do relate the variation in a figure of merit (FOM) to the variations in parameters. Very often tolerance is the word used to describe the total variation in the FOM due to all of the variations in the parameters, and sensitivity is reserved for a measure of the change in the FOM compared to a change in a particular parameter, the other ones remaining constant.

All of these measures have a disadvantage inherent in them, viz., they do not convey all of the information about a system that there is to convey. The more numbers one uses the more information he can convey, but the more difficult it is to assimilate; the fewer numbers one uses, the less information he can convey, but the easier it is to assimilate. It is for this reason that the multitude of definitions has arisen.

To discuss sensitivity and tolerance precisely definitions are required. First suppose that there is an FOM (e.g., gain of an amplifier) expressed in terms of a set of independent parameters ($\underline{x}$); this is the $y = g(\underline{x})$ formulation mentioned in Sec. 5. Then sensitivity of the FOM to each parameter is defined here* as

$$S_i \equiv \left. \frac{\partial \text{ FOM}}{\partial x_i} \right|_{\underline{x}}$$

where the $\underline{x}$ refers to the entire set of components (x_1, x_2, ...)
and the $|\underline{x}$ refers to the variables considered independent for this differentiation. (See Appendix D for details on tricky partial differentiation.)

Next suppose that the situation is more complicated. Let there be an equation for the FOM in terms of the parameters, $\underline{x}$, and several equations relating the x_i to each other and possibly to other parameters. Let $\underline{u}$ be the subset of parameters of $\underline{x}$

---

* The definitions of sensitivity to follow are selected from the possible ones given in Sec. 5.1.

that are to be considered independent for this particular case. Then the sensitivity of the FOM to the parameter u_i can be defined as

$$S_{i,\underline{u}} \equiv \left.\frac{\partial\ \text{FOM}}{\partial\ u_i}\right|_{\underline{u}} = \sum_j \left.\frac{\partial\ \text{FOM}}{\partial x_j}\right|_{\underline{x}} \times \left.\frac{\partial x_j}{\partial u_i}\right|_{\underline{u}} .$$

There is no law which determines which subset of all the parameters is chosen to be independent. It is up to the engineer to pick the one to give the most useful sensitivity. Once in a while one may wish to calculate an $S_{i,\underline{u}}$ and an $S_{i,\underline{v}}$ where $\underline{u}$ and $\underline{v}$ are different subsets of the total set of parameters; it is a reasonable thing to do.

A relative sensitivity is defined in Sec. 5.1. When it is adapted to the situation described above, the relative sensitivity of the FOM to the parameter u_i (with the set $\underline{u}$ taken as independent parameters) becomes

$$S^*_{i,\underline{u}} = \frac{u_i}{\text{FOM}} \times \sum_j \left.\frac{\partial\ \text{FOM}}{\partial x_j}\right|_{\underline{x}} \times \left.\frac{\partial x_j}{\partial u_i}\right|_{\underline{u}} \equiv \sum_j \left.\frac{\partial\ \ln\ \text{FOM}}{\partial\ \ln\ x_j}\right|_{\underline{x}} \times \left.\frac{\partial\ \ln\ x_j}{\partial\ \ln\ u_i}\right|_{\underline{u}} .$$

This latter could be interpreted in an example as: There will be a 1% change in the FOM, when u_i changes by 10% and the rest of the $\underline{u}$ remain constant.

Other sensitivities, depending on one's needs, can be similarly defined. Even though the notation is complicated at first (and second) glance, it can be deciphered by following it through. For simple examples, see Sec. 5.1. The reason for the complexity is that it is so easy to become hopelessly snarled during a partial differentiation, unless one is quite rigorous with the notation.

Tolerance is usually defined as the variation in an FOM due to a combination of variations in the independent parameters. Before this has much meaning it must be decided how these will combine. Two simple methods of combining are worst-case and "statistical". The worst-case tolerance (for independent parameters, $\underline{u}$) is

$$\Delta\ \text{FOM}\Big|_{wc,\underline{u}} \equiv \sum_i \left|S_{i,\underline{u}}\ \Delta u_i\right| \equiv \sum_{ij} \left|\left.\frac{\partial\ \text{FOM}}{\partial\ x_j}\right|_{\underline{x}} \times \left.\frac{\partial\ x_j}{\partial\ u_i}\right|_{\underline{u}} \times \Delta\ u_i\right| \ *\ .$$

The statistical tolerance usually means the standard deviation of the FOM. Except for a very few cases, the only function for which the standard deviation is easily calculated is a linear function. A Taylor's series expansion (see Sec. 5.1.2

---

* | | means the absolute value of what's inside the parallel bars.

and Appendix B) is most often used to make a function linear. Then

$$\sigma^2(\text{FOM}) \equiv \sum_i (S_{i,\underline{u}})^2 \Big|_{\underline{x}_o} \sigma^2(u_i) \equiv \sum_i \left(\sum_j \frac{\partial\ \text{FOM}}{\partial\ x_j} \Big|_{\underline{x}} \times \frac{\partial\ x_j}{\partial\ u_i} \Big|_{\underline{u}} \right)^2 \Big|_{\underline{x}_o} \sigma^2(u_i)$$

where $\underline{x}_o$ is the point about which the Taylor's series is taken (and is the mean of the linear function--about which the variance is taken). This formula, as Sec. 5.2.1 shows, is for statistically independent variables--which is what the subset $\underline{u}$ has been presumed to be. This formula is true regardless of the pdf's of the $\underline{u}$ and FOM, but remember--there are no probability statements associated with it. In particular, avoid using Normal (Gaussian) probabilities. The Central Limit theorem is often invoked to show that the FOM has a Normal distribution, but this may be inadequate, as is explained in Appendix E, as well as because the calculation is based on dropping all but the linear terms in a Taylor's series expansion.

For some kinds of functions there are special relationships between all the sensitivities for a given function (see Ref. 7). This can be a help in both doing and checking the analysis.

6.2 Worst-Case Analyses

One of the apocryphal laws of nature is "if it can happen, it will". It is given different names by various groups and has possibilities of being extended, e.g., "if it can happen, it not only will, but it will in the most embarrassing way possible". Much reliability effort is devoted to reducing the adverse effects of this law on the life and performance of equipment. It was natural then that something called "worst-case" design and analysis would spring up in reliability; no attempt is made here to define the concept of worst-case design as opposed to analysis, in fact, it is difficult to do.

At this point the plural in the section title becomes important because it turns out that different people have different sets of specific criteria for worst-case. Obviously the worst worst-case is where everything falls apart or otherwise ceases to function, and just as obviously nothing can work under those circumstances. So something else is meant by worst-case, viz., a set of limits is established by some criteria (to be discussed later) and within those limits a worst-case analysis is to be performed. For example, if it is assumed that the resistors will not deviate by more than 10% for any reason during the design life of the equipment, will all the circuits continue to function, even if the resistors get spiteful about it? In this connection, probabilities except for zero and one are ignored, but direct physical dependence should be taken into account. For example, if a voltage bus feeds several different points, the voltages at each of the several points should not be treated as

variables independent from each other. Likewise if temperature coefficients are taken into account, one part of the system should not be presumed to be at the hot limit and the other at the cold limit at the same time--unless of course it is physically reasonable that it be so. It is presumed in the following discussion that these correlations of zero or one are taken into account wherever possible. A very general boundary condition for the analysis is that the circuit or system should be constructed according to the specifications in the drawings and that the analysis proceeds from there. The fact that there may be a nonzero probability that some parameters will fall outside the limits established for the analysis is ignored. The most likely candidate for setting the limits referred to above is absolute worst-case (AWC). In AWC the limits for each independent parameter are set without regard to other parameters or to its importance in the system. The position of the limits is usually set by guess or by golly; the engineer must estimate these by reference to his experience, to the manufacturer's data, and to his company's experience (see Sec. 7). In some cases he will perform several analyses with different limits for each--just to get a feel for the situation. Obviously, the entire procedure of setting limits has no strict rules.

There are many modified worst-case (MWC) analyses developed because of the presumed pessimism of AWC. It is not worthwhile going into all of these here, but a typical one uses the following method for setting the limits: Critical items are given limits as in AWC and the rest of the items are given limits of their purchase tolerance.

In any worst-case analysis, the values of the parameters are adjusted (within the limits) so that the FOM is as high as possible, then readjusted so it is as low as possible. The values of the parameters are not necessarily set at the limits--the criterion for their value is to make the FOM an extreme. The remoteness of this condition's occurring in practice depends on the limits which were set by the engineer at the beginning, on the probability functions of the parameters, and on the number of systems that are being considered.

One argument in favor of AWC analysis (as opposed to a statistical analysis) is that many digital electronic systems have so many similar parts, each of which must have such a high probability of working properly, that a statistical analysis will for practical purposes turn out to be an AWC analysis, and the AWC is much simpler and depends on fewer tricky assumptions. The way to make a useful AWC analysis is to set the limits wisely in the first place and it is here that engineering judgment versus statistical calculation must come to the fore. Where the importance warrants it and time and money permit, several such limits should be set up and the calculations made for each.

The important thing in the analysis is not naming it (as long as the name is not misleading) but understanding what you are doing and why. It may be that the reasons are to satisfy a recipe prescribed in the specification, in which case performing the calculation by rote is sufficient. But generally the designer will have to ask himself, "what is it that I would like to have?", then go ahead and make the calculations. If all looks reasonable, he can be happy; if it doesn't look reasonable, he will either have to modify his design, or modify what he would like to have (sometimes the choice between the latter two is not always his).

Computer routines are available for performing these analyses on electronic circuits. Generally speaking the curve of FOM versus each independent parameter is assumed to be monotonic and a numerical differentiation is performed at the nominal values to see in which direction the parameter should be moved to make the FOM high or low. It is also presumed that this direction is independent of the values of any of the other parameters as long as they are within their limits. If these assumptions are not true, a much more detailed analysis of the equations is necessary before worst-case can be performed. Essentially what it would amount to is that a response surface for the FOM must be generated for all the parameters involved. It is seldom that anyone will feel such a complicated analysis is worthwhile.

An interesting modification of worst-case is given in Ref. 8, it is called "Worst Distribution Analysis". This is less stringent than the worst case above and has promise of being an easy, effective calculation.

6.3 Propagation of Probability Distributions Through an Equation

In this section it is presumed that an equation is available in the form $y = g(\underline{x})$ and that the probability density functions (pdf) of all the x_i are known exactly. The most usual parameter that engineers want (perhaps because it is one with which they are most familiar) is an average or mean. An engineer who has been exposed to the vagaries of the world (as opposed to the uniformity of textbooks) soon learns that there is more of interest than an average, that the variations are important, and therefore he wishes to calculate the standard deviation or variance (a second moment). But even in this situation, for the first two moments,* there are not many tractable

---

* There are many measures of "central tendency", the mean (the "center of gravity" of the population), the median (half the population lies on either side of the median--without regard to how big or small the values are), and the peak (called a "mode" by statisticians--it is the most frequently occurring value, e.g., in the Normal distribution it is the mean; in the exponential distribution it is zero!). These measures are useful only when the distribution has one peak and is not too cockeyed (i.e., skewed). If it has more than one peak, it is often desirable to try to split it into two parts, each of which has only one peak (i.e., unimodal).

The variance corresponds to the physical moment of inertia about the center of gravity.

combinations of equations and probability-functions. So most often one will have to resort to approximate methods such as discussed in Sec. 5. Formulas are given in Appendix F for some tractable sets of equations and probability distributions.

In the more general case pdf(y) is desired; but rarely can any system equation be handled exactly. Again reference should be made to Appendix F for a listing of some of the tractable combinations. The opening statement of this section tends to imply that all x_i are statistically independent; otherwise each would not have its own pdf. If there are statistical dependencies among some of the variables, the problem becomes very difficult unless the variables are Normal or the combined pdf($\underline{x}$) is known. The services of a statistician will be virtually indispensible in the event that statistical dependence must be considered.

Assuming that the variables are statistically independent it is easy enough to write down the integral according to the formula in Sec. 5. As mentioned there, in the absence of a fortuitous circumstance wherein the result is tractable (the engineer sometimes can make his own luck by choosing the appropriate distributions for some of the variables), the equation cannot be analytically integrated and numerical methods must be used. One usually resorts to the use of computers. Often the engineer will not possess the required knowledge of numerical methods and will have to seek help on that score. He may even have to get assistance on the programming unless he has access to a computer with special programs designed to make programming extremely simple.

If ordinary numerical integration does not seem appropriate or feasible, Monte Carlo methods can be used. They are effective, but often are expensive as well since they tend to consume a great amount of computer time. If only the very central region of the pdf is of interest, perhaps as few as 20 or 30 trials will be feasible. Some general rules of thumb are given in Sec. 5 for determining approximately how many runs should be made.

An alternate method which is sometimes feasible where the moments of $g(\underline{x})$ can be calculated is to compute successive moments starting with the first and then use standard statistical methods for fitting one of the appropriate distributions to these moments. It is difficult to evaluate how far in the tail region of the pdf these will be good, but the odds that they are accurate for tail areas less than 1% are not good; the chances of being accurate for tail areas less than 0.1% are negligible.

If the pdf (x_i) are estimated from small samples, the techniques in the Sec. 6.4 are applicable.

6.4 Estimation of Parameters of a Distribution

This section is more general than the title might appear since the parameters of the distributions can be involved in equations and it is the parameters of these equations that will eventually be estimated by the method. One of the most common examples is estimating the slope and intercept of a straight line by using the method of Least Squares. Typical methods which are used are Maximum Likelihood, Least Squares, Equating of Moments, and Order Statistics. These are detailed in various statistics texts and general articles and will be only summarized here.

Maximum Likelihood. In this method the probability (for discrete variables) and pdf's (for continuous variables) must be known for the variables, assuming that the parameters of the distributions are given.* It is generally presumed that all observations are statistically independent so that the probability of getting a given set of observations is the product of the probabilities of getting each observation. The total probability is called the Likelihood. If the parameters of the distributions must obey some equations, these equations are then inserted either directly or used as constraints. The ultimate set of parameters is then adjusted so that the Likelihood expression, written previously, is a maximum. At this point, equations are easier than words. Suppose that the observations are denoted by Obs, and the unknown parameters by $\alpha,\beta, \ldots,\gamma$. Then L is a function of Obs and $\alpha,\beta, \ldots,\gamma$, i.e., $L = L(\text{Obs},\alpha,\beta, \ldots,\gamma)$. The customary method of finding a maximum is followed, viz., a set of values of $\alpha,\beta, \ldots,\gamma$ is found, designated by $\hat{\alpha},\hat{\beta}, \ldots,\hat{\gamma}$,** such that

$$0 = \left.\frac{\partial L}{\partial \alpha}\right|_{\wedge} = \left.\frac{\partial L}{\partial \beta}\right|_{\wedge} = \ldots = \left.\frac{\partial L}{\partial \gamma}\right|_{\wedge}$$

where $|_{\wedge}$ means evaluated at the $\wedge$ conditions.

It is often desirable to have the estimates of the parameters as independent of each other as possible. This can be done (asymptotically, i.e., the larger the sample the truer it is) by adjusting the formulation of the equations (e.g., definitions of $\alpha,\beta, \ldots,\gamma$ and the origins of the independent variables) so that

$$0 = \left.\frac{\partial^2 L}{\partial \alpha \partial \beta}\right|_{\wedge} = \left.\frac{\partial^2 L}{\partial \alpha \partial \gamma}\right|_{\wedge} = \ldots = \left.\frac{\partial^2 L}{\partial \beta \partial \gamma}\right|_{\wedge} ,$$

i.e., all the "mixed" second derivatives (parameters only) are zero. It should be

---

* This is the form in which they usually appear, e.g., the Poisson probability is $p(n|\mu) = e^{-\mu}\,\mu^n/n!$. The expression "$p(n|\mu)$" is read "the probability of n, given that μ is known."

** The $\wedge$ is called a "circumflex" or simply a "hat".

noted that virtually always it is more convenient to deal with the natural log of the Likelihood rather than the Likelihood itself. If that is done, all of the products are converted to sums and differentiation is much easier. Furthermore, it is easy to manipulate the equation so that the constant terms (terms which do not contain the $\alpha, \beta, \ldots, \gamma$) are combined with the ln L. They will drop out for all differentiations anyway and to get rid of them in the beginning is a good way to keep the problem simple. (That means you won't make as many stupid mistakes in your work.)

Note the following:

(1) We will maximize L.

(2) (1) is equivalent to maximizing ln L.

(3) (2) is equivalent to maximizing $-\ell = \ln(L \times \text{constants}) \equiv \ln L + \text{constants}$.

(4) (3) is equivalent to minimizing ℓ.

The reason for introducing ℓ, defined in this way is for the simplicity above and to ease the introduction of the estimated variance of the parameters. It is instructive to expand ℓ about the $\hat{}$ point in a Taylor's series:

$$\Delta\ell = \left.\frac{\partial \ell}{\partial \alpha}\right|_{\hat{}} \Delta\alpha + \left.\frac{\partial \ell}{\partial \beta}\right|_{\hat{}} \Delta\beta + \ldots + \left.\frac{\partial \ell}{\partial \gamma}\right|_{\hat{}} \Delta\gamma$$

$$+ \left.\frac{\partial^2 \ell}{\partial \alpha \partial \beta}\right|_{\hat{}} \Delta\alpha\Delta\beta + \left.\frac{\partial^2 \ell}{\partial \alpha \partial \gamma}\right|_{\hat{}} \Delta\alpha\Delta\gamma + \ldots + \left.\frac{\partial^2 \ell}{\partial \beta \partial \gamma}\right|_{\hat{}} \Delta\beta\Delta\gamma$$

$$+ \tfrac{1}{2} \left.\frac{\partial^2 \ell}{\partial \alpha^2}\right|_{\hat{}} \overline{\Delta\alpha}^2 + \tfrac{1}{2} \left.\frac{\partial^2 \ell}{\partial \beta^2}\right|_{\hat{}} \overline{\Delta\beta}^2 + \ldots + \tfrac{1}{2} \left.\frac{\partial^2 \ell}{\partial \gamma^2}\right|_{\hat{}} \overline{\Delta\gamma}^2$$

$$+ \text{higher order terms}$$

The terms in first line are each zero by definition of the $\hat{}$ point. The terms in the second line are zero because of the orthogonality condition. The fourth and last line is presumed negligible. Therefore

$$\Delta\ell \simeq \tfrac{1}{2} \left.\frac{\partial^2 \ell}{\partial \alpha^2}\right|_{\hat{}} \overline{\Delta\alpha}^2 + \ldots + \tfrac{1}{2} \left.\frac{\partial^2 \ell}{\partial \gamma^2}\right|_{\hat{}} \overline{\Delta\gamma}^2,$$

and each parameter can be considered by itself. It turns out that, for each parameter by itself, when $\Delta\ell = \tfrac{1}{2}$, $\overline{\Delta\alpha}^2$ = est var α, ..., $\overline{\Delta\gamma}^2$ = est var $\hat{\gamma}$.* (This is explained in detail in Ref. 9 along with other properties of maximum likelihood.)

---

* est var $\hat{\alpha}$ means the estimated variance of $\hat{\alpha}$, etc.

Therefore

$$\text{est var } \hat{\alpha} = \left(\frac{\partial^2 \ell}{\partial \alpha^2}\bigg|_{\hat{}}\right)^{-1}, \ \ldots, \ \text{est var } \hat{\gamma} = \left(\frac{\partial^2 \ell}{\partial \gamma^2}\bigg|_{\hat{}}\right)^{-1}$$

and the uncertainty in a parameter estimate can be taken as the square root of the estimated variance. The estimates of the variances of the parameter estimates are asympotically correct.

Example of Maximum Likelihood. Suppose that

$$y = g(x) + \sigma z$$

where $g_i \equiv g(x_i)$, $\sigma_i \equiv \sigma(x_i)$, $z_i \equiv \dfrac{y_i - g_i}{\sigma_i}$

(x,y) are a set of independent, dependent variables as is customary

z has a Standard Normal distribution

σ^2 is the variance of y.

(x_i, y_i) is a set of corresponding data. All sets are statistically independent. Then the probability of z's being between z_i and $z_i + dz_i$ is

$$p(z_i)dz_i = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2} dz_i$$

If there are n point sets, the likelihood L is just the product of the probabilities for each of these point sets (statistical independence was assumed--see Sec. 2.10):

$$L = p(z_1)dz_1 \times p(z_2)dz_2 \times \ldots \times p(z_n)df_n \equiv \prod_{1i}^{n} p(z_i)dz_i$$

$$= \prod_{1}^{n} e^{-\frac{1}{2}z_i^2} \frac{dz_i}{\sqrt{2\pi}}$$

since z is Standard Normal variate.

The x_i and thus σ_i and g_i are presumed fixed and exact, so that $dy_i = \sigma_i dz_i$. The dy_i are constants and are tucked out of the way as mentioned above, viz.,

$$\ell = -\ln\left(L/\prod_{1i}^{n} \frac{dy_i}{\sqrt{2\pi}}\right) = \frac{1}{2} \sum (z_i^2 + \ln \sigma_i^2) = \frac{1}{2} \sum \left(\frac{(y_i - g_i)^2}{\sigma_i^2} + \ln \sigma_i^2\right).^*$$

We wished to maximize L, therefore we minimize $\ell = -\ln(L/\text{constant})$. If $\sigma_i = \sigma$ is a constant then minimizing ℓ is equivalent to minimizing

---

* $\sum$ implies $\sum_{1}^{n}{}_i$

$\sum (y_i - g_i)^2$ which is exactly the usual formulation of least-squares. If $\sigma_i \neq$ constant, note that each square is weighted by $1/\sigma_i^2$ (the reciprocal of the variance) and that there is a logarithim of the variance added.

Still considering maximum likelihood it is instructive to derive $\hat{\sigma}$ (assuming σ is a constant).

$$\ell = \frac{1}{2} \sigma^2 \sum (y_i - g_i)^2 + n \ln \sigma ,$$

$$\frac{\partial \ell}{\partial \sigma} = - \frac{1}{\sigma^3} \sum (y_i - g_i)^2 + \frac{n}{\sigma} ,$$

$$\left.\frac{\partial \ell}{\partial \sigma}\right|_{\hat{}} = 0, \qquad \hat{\sigma}^2 = \frac{1}{n} \sum (y_i - \hat{g}_i)^2 .^*$$

This shows that the Maximum Likelihood estimate for the variance of a Gaussian distribution is just the sample variance, if σ is a constant. If the data have been transformed, it is very unlikely that σ is a constant.

Least Squares. This technique has merit in its own right regardless of its connections with Maximum Likelihood and the Normal distribution. From the example for Maximum Likelihood above,

$$\ell = \frac{1}{2} \sum \left(\frac{(y_i - g_i)^2}{\sigma_i^2} + \ln \sigma_i^2 \right) . \qquad \text{(general case)}$$

If $\sigma_i = \sigma =$ constant, then (also as in the example above)

$$\ell_1 = \frac{1}{\sigma^2} \sum (y_i - g_i)^2 + n \ln \sigma^2 . \qquad \text{(special case)}$$

As far as the parameters in g_i are concerned, minimizing ℓ is equivalent to minimizing $\sum (y_i - g_i)^2$--whence the name "Least Squares". In general σ is not a constant and ℓ (general case) is the quantity to be minimized, but the name of Least Squares is still used. The estimates have some very good properties under special circumstances of the parameters, regardless of Normality of $y - g(x)$.** For distributions such as the exponential which are highly skewed, the method has little value. If $g(x)$ is not linear or $\sigma = \sigma(x)$, the problem is not likely to be tractable. Then numerical methods must be used to find the parameters.

---

* $\hat{g}_i$ implies that all parameters in g_i are evaluated at the $\hat{}$ condition.

** See Ref. 10 for further details on Least Squares.

Example a. Let $g(x) = \mu$ = constant, $\sigma(x) = \sigma$ = constant. This is the situation where a constant is estimated by a series of observations.

$$\ell = \frac{1}{2} \sum \left(\frac{(y_i - \mu)^2}{\sigma^2} + 2 \ln \sigma \right)$$

$$0 = \left.\frac{\partial \ell}{\partial \sigma}\right|_{\wedge} = -\frac{1}{\hat{\sigma}^3} \sum (y_i - \hat{\mu})^2 + \sum \frac{1}{\hat{\sigma}} , \quad \text{or } \hat{\sigma}^2 = \frac{1}{n} \sum (y_i - \hat{\mu})^2$$

$$0 = \left.\frac{\partial \ell}{\partial \mu}\right|_{\wedge} = \frac{1}{\hat{\sigma}^2} \sum (y_i - \hat{\mu})(-1) , \quad \text{or } \hat{\mu} = \frac{1}{n} \sum y_i$$

$$\left.\frac{\partial^2 \ell}{\partial \mu \partial \sigma}\right|_{\wedge} = 0 , \quad \left.\frac{\partial^2 \ell}{\partial \mu^2}\right|_{\wedge} = \frac{n}{\hat{\sigma}^2} , \quad \left.\frac{\partial^2 \ell}{\partial \sigma^2}\right|_{\wedge} = \frac{2n}{\hat{\sigma}^2}$$

est var $(\hat{\mu}) = \frac{\hat{\sigma}^2}{n}$, est var $(\hat{\sigma}) = \frac{\hat{\sigma}^2}{2n}$,where "est var" stands for "estimated variance of". This says that the least squares estimate of a group of data which measure a constant is the sample average and that the standard deviation (s.d.) estimate is the sample s.d. Note especially that $\hat{\sigma}^2$ is biased. It can be converted, if it is worth it, to an unbiased estimate by the usual techniques (see Sec. 2.11 for a discussion of this point).

Example b. Let $g(x) = mx$, $\sigma(x) = kx$. This is a straight line through the origin with a constant "percentage-accuracy" for y, viz., $\sigma(x)/g(x) = k/m$.

$$\ell = \frac{1}{2} \sum \frac{(y_i - mx_i)^2}{k^2 x_i^2} + \sum \ln k\, x_i$$

Let $\left.\frac{\partial \ell}{\partial m}\right|_{\wedge} = \left.\frac{\partial \ell}{\partial k}\right|_{\wedge} = 0$. Evaluate $\left.\frac{\partial^2 \ell}{\partial m \partial k}\right|_{\wedge}$ and note that it is zero; therefore the estimates of m and k are asymptotically independent.

It is easily shown that

$$\hat{m} = \frac{1}{n} \sum \frac{y_i}{x_i} , \quad \hat{k} = \frac{1}{n} \left\{ \sum \left(\frac{y_i}{x_i}\right)^2 - n \left(\sum \frac{y_i}{x_i} \right)^2 \right\}^{\frac{1}{2}}$$

$$\text{est var } \hat{m} = \hat{k}^2/n, \quad \text{est var } \hat{k} = k^2/2n .$$

Equating of Moments. The sample moments are calculated beginning with the first. The number so calculated should be the same as the number of unknown parameters in the distribution. The analytic expressions for the moments of the distribution are also calculated. Corresponding sample and distribution moments

are equated and the equations are solved for the parameter estimates. Usually the equations are not tractable and complicated numerical methods for solving all equations simultaneously must be used.

The Pearson System of curves which contain several arbitrary parameters are well suited to the Equations of Moments. Ref. 11 contains a brief discussion of the actual procedures and gives references to more complete treatments.

It is difficult to estimate the uncertainties which are incorporated into the answers by the Equating of Moments and so this technique is best left alone when samples are small and the uncertainties will be large.

Order Statistics. This technique is especially valuable when observations are left off of either or both ends. (Truncation and censoring are technical words used to describe these omissions.) There is no general method for applying it since simulation seems to be one of the main ways in which the necessary data are generated for use in forming the rules. Ref.12 is reasonably up to date, but the current literature--including Government Reports--contains much useful information. A statistician who has this as a specialty will be a big help in using it.

At times, this method is mixed with others, especially regarding the estimation of the "guarantee period"* for some distributions. For example, the smallest observation is an estimate of the guarantee period and is not infrequently used as such.

When no simple distribution is assumed, the cumulative distribution can be estimated by equating the sample cumulative distribution to the actual one at the sample points. This is where plotting position comes in (for the i-th point what probability should be used?). An easy value to use for the i-th point out of n is i/n+1; this is the expected value. There are many others offered in the literature, but don't forget this admonition:

> The uncertainties involved in the prediction are usually much greater than differences among plotting positions.

Therefore if you think the exact formula for plotting position is critical, you have problems which will not be solved by choosing a better plotting position (in other words, your troubles are for psychiatrists, not statisticians.) The figures in Sec. 5.4 show the variations which can easily happen.

Graphing. One of the common engineering methods for estimating a probability distribution is to plot the cumulative distribution function (cdf) on special graph paper such that the analytic cdf is a straight line. Weibull paper, semi-log paper (for the exponential) and Normal** paper are the only ones in common usage. Normal

---

* See Appendix G.

** Often Normal paper is called Probability paper as if it were the only kind that existed; that is poor practice.

paper is available with a $\log_{10}$ scale for the other coordinate for use with log Normal distributions. The figures in Sec. 5.4 show the scatter that readily occurs with this method; so for small samples, plotting as a straight or curved line doesn't prove anything. This method does not allow the easy estimation of uncertainties and so is disadvantaged.

It is tempting to calculate by means of Least Squares the equation for this line--Resist It! Least Squares requires statistically independent points. The ones on a cdf are not statistically independent, because they are ordered. That is, given the second point, you know something about the third one, viz., it is above the second. Since one point gave you knowledge about the other, they cannot be statistically independent (see Sec. 2.10). Some work has been done with order statistics to overcome this problem, but the results are not generally applicable.

Goodness of Fit. After fitting a distribution to a set of data, it is customary to see how good the fit is. First of all, you have to decide on criteria for goodness of fit. Are you going to use an engineering or statistical criterion? The big difficulty with statistical criteria is their very nature. The question a statistical test answers is: What are the chances of picking some sample from the assumed distribution and having that sample more oddball than your sample? If there are only a few points, the chances are pretty good that any sample is oddball (see figures in Sec. 5.4). This is known as the test's having low discriminating ability (called "power" in statistics). If there are a great many points, the chances are rare that any sample from the assumed distribution will be as oddball as your actual data. In this case, the test has too high a discriminating ability. But the question the engineer usually wants to have answered is: If I go ahead and assume a distribution, will I be too far off in my calculations? And this question can't be directly answered. Calculation of uncertainties for parameter estimates, and consequent uncertainties in calculations using them, are one way of putting a lower bound on the possible errors. Plotting the cdf on appropriate paper (so the cdf is a straight line) can also show a lower bound on potential errors. If getting the proper distribution is quite critical, it pays to try several and to see what effect the different ones have on the prediction.

Two statistical tests in common use are Chi-Square and Kolmogorov-Smirnoff. The Chi-Square test compares the actual number of observations in an interval with the predicted number. It combines these comparisons for all intervals and the resulting statistic has a Chi-Square distribution with parameter somewhat less than the number of intervals. It is explained in detail in many statistics texts, but generally has low discriminating ability for reasonable sample sizes.

The Kolmogorov-Smirnoff test checks on the maximum deviation of the actual cdf from the predicted cdf. The tables for use with it are distribution-free; so the test lends itself well to graphical methods. For that reason it is becoming more widely used by engineers. It is explained and referenced further in Ref. 13.

There are other statistical tests, some are peculiar to a distribution, others are distribution free. A statistician should be consulted if you want to use these tests and don't want to become a statistician yourself. Just be sure you have him explain the principles involved in the test (you need not be concerned about the details) so that you can judge for yourself what the test answer really tells you.

Difficulties with Discrete Distributions. If the random variable is discrete, such as in the Poisson and binomial distributions, estimation of confidence limits becomes tricky. You can estimate upper and lower bounds for each end point of the confidence interval. (See Ref.11 for a further discussion of this point) or you can ignore the problem (it won't go away, but you can ignore it) or you can find a statistician who has this area as a specialty and find come compromise estimates of the end points of the confidence interval.

6.5 Calculation Where Extreme Extrapolation Is Necessary.

One of the main difficulties with the calculations involved in Secs. 6.3 and 6.4 above for high reliability purposes is that the region where the calculations are most certain is the one of least interest to the reliability engineer. The region of relative certainty, usually near the center of the data, is readily estimated and handled, and the problems readily corrected. The importance of this central body should not be denigrated, but since it is soon taken care of, the reliability engineer's attention is directed toward the regions where there are few data, where the uncertainties are high, and where the probabilities are low.

It is easy at this point to be dazzled by the accuracy of numerical calculations. For example, if a Weibull distribution is fitted to the data, it can be extrapolated far, far out into the tail region beyond any data, and calculations can be precisely made with as many significant figures as tables or computer routines are available. But this result only tells you what would happen if in fact the Weibull were an accurate description of the tail region; it does not tell you that the Weibull distribution is in fact such an accurate description. Most likely, none of the tractable common distributions are descriptive in the tail regions. That is unfortunate, but it is nevertheless true. At this point Sec. 5.4 should be consulted for the unpleasant choices which face the engineer under these circumstances.

7. Potential Sources and Uses of Variation Data

Here is where the difficulties become even more apparent: no matter whether the items are purchased or made in-house, there won't be enough data on them.

7.1 Purchased Items

7.1.1 Manufacturer's Data

These should rarely be relied upon in critical applications unless incoming-inspection tests confirm the data. For example, not all manufacturers check all devices to see that they fall within the maximum and minimum specifications. Distributional data are to be particularly suspect because they rarely account explicitly for:

(1) distribution within a lot,

(2) day-to-day variations, and

(3) variations among different lots.

These sources of variability create a much broader distribution for periods of weeks, months, or years than one gets in a single lot. In addition, especially if the item is stock and standard, the manufacturer may sort the lot prior to its being shipped to you in order to extract from it a sublot meeting the requirement of some other customer. For example resistors which have a nominal 10% tolerance may have a large portion of the central region subtracted from them to meet the requirements of those who wish a 5% resistor. The resulting distribution could easily have a double peak and a minimum around zero deviation from nominal.

The published data on properties of metals are in worse shape than those of electronics. Those who try to reconstruct distributions of strengths or other properties from one or two points (and these points have only nominal probabilities) do so at their own peril. A good source for and discussion of metal properties is Ref.14 and associated material issued later.

If the distributional data of the manufacturer are critical and must be relied upon, it is wise to have an express agreement to that effect with the manufacturer. Provision must be made for enough sampling of important parameters by either the manufacturer or the customer to assure conformance.

The state-of-the-art with regard to quality of elements seems to depend on whether a producer or a consumer of the element under question is doing the discussing. Producers are naturally much more optimistic; consumers tend to be somewhat pessimistic. It is especially unwise to take the manufacturer's maximum and minimum limits and presume that they are $\pm 3\sigma$ limits from a Normal (Gaussian) distribution. For mechanical and electromechanical items such as pumps, motors, and relays the same cautions hold, especially if one is pushing the state-of-the-art. Anything

that is difficult to make is much more likely to be made wrong. What we mean by state-of-the-art is those elements which are on the borderline between being easy and difficult to produce.

7.1.2 Acceptance and Receiving Tests

Historical data from acceptance and receiving tests will show how the parameters of the elements have been behaving in the past. While it is common practice to extrapolate from past to future performance, a close watch should be kept on the data so that violent changes will be signaled as soon as possible. The manufacturer may shift his process for his own reasons such as product improvement, process simplification, greater profit, or longer life or it may shift without his knowing it. Sometimes when he considers he has improved his product, he may have changed it for the worse for you.

It is common practice to fit tractable distributions such as the Normal directly to parameter data or to data which have been transformed so that they will fit better. The logarithm of a datum is a common transformation. Since analyses for reliability are often concerned with the behavior of these distributions way out in the tails, it is the fit of a distribution out on the tails that is important. But this is the region of the curve that is difficult if not impossible to fit from the data. Just because the data fit a Normal distribution around the middle is no reason at all to feel that they fit it in the tails. There may be considerable skewing to the right or to the left and there may be large bumps in the tail regions. Any of these can cause predictions in the tails to be off by factors from 10 to 10^4.

These acceptance and receiving tests may give good information on the reversible effects such as temperature and supply voltage, and can be quite useful in parameter variations analysis.

7.1.3 Special In-house Tests

Part qualification tests usually assume that variability in parts parameters (and FOM's) is negligible and typically rely on small sample sizes. These tests can be useful for estimating the magnitude of reversible effects. Also some elements can be intentionally varied during the qualification tests and the effect on the system parameters observed directly.

7.1.4 Coordinated Data Facilities

Facilities such as IDEP* usually provide extensive amounts of raw data. But since test conditions vary so widely and so many circumstances are unknown or

---

* See Volume V--Parts for a more extensive discussion of these facilities.

different from those desired, the raw data can be used only as a rough guide to the variability of parameters of the elements. In particular, since you don't know who ran the test, you don't know how well he did it. Some databank facilities are being restructured; others being built anew to overcome these difficulties; and still others are dying.

7.1.5 Field Data for Operating Systems

Field data range from excellent to worthless depending on the system and the people who make it work. Field data will generally be suitable more for qualitative than quantitative use.

7.1.6 What to Do

Often the engineer will have to use arbitrary distributions or set arbitrary limits when making analyses and use the information from the above sources as a guide. His answers will then be of the form: "If the element parameters vary in such a way, then the system parameter will vary thusly." It is up to the reader then to see if he thinks the assumptions were reasonably realistic. Even though this latter approach is often not explicitly followed it is in effect the practical outcome of many of the calculations. When it's the best you can do, there's nothing at all wrong with it.

7.2 Manufactured Items

The major data sources for manufactured items are in-house testing and field experience data. These are as inclusive as the resources of the company permit and the attitudes of the company allow. They should not be overlooked. Many of the same considerations hold as for purchased items, especially when a different plant in the company makes them or when the department that makes them has different goals than the user does. One can as readily be done-in by someone inside his company as by someone outside it.

8. Summary and Conclusions

Once the concepts involved in parameter variations analysis are understood, the application becomes one of an engineer's deciding what techniques will best serve his needs. Without understanding the concepts involved, the effective use of this kind of analysis is virtually impossible.

There are many occasions where the full treatment is not feasible because of budget restraints or because the system requirements are reasonably loose. Before embarking on an extensive program of parameter variations analysis it is wise for the engineer to spend a few hours being explicit about the advantages he will get if the program is successful. This will help in evaluating the desirability and necessity of such a program. In virtually all cases, try the quick-and-dirty methods first and go on to the more complicated ones if they are necessary, rather than jumping into the more complicated ones right away.

Much of the terminology in the field can be confusing and can lead engineers to think that they do not understand what needs to be done; whereas virtually all of the tools are fairly simple. The proliferation of names is unfortunate from this point of veiw even if understandable from an individual author's point of view.

Perhaps the two most important subsections in this volume are 2.1 (the concept of a model) and 5.4 (techniques for analyzing mathematical models where extreme extrapolation is necessary). These are most important because

(1) they are least likely to be understood

(2) this lack of understanding is most likely to cause trouble.

References

1. Bhattacharya, S. K.: Bayesian Approach to Life Testing and Reliability Estimation. JASA 62, pp. 48-62.

2. Freund, J. E.: Mathematical Statistics. Prentice Hall, 1962.

3. Abramowitz, M.; and Stegun, I. A., eds.: Handbook of Mathematical Functions. AMS 55. Fifth printing (with corrections). U. S. Government Printing Office, August 1966.

4. Gradshteyn, I. S.; and Ryzhik, I. M.: Table of Integrals, Series, and Products. Fourth ed., Translated from the Russian. Academic Press, 1965.

5. Richards, P. I.: Manual of Mathematical Physics. Pergamon Press, 1959.

6. Franklin, P.: Treatise on Advanced Calculus. John Wiley & Sons, Inc., 1940.

7. Belove, C.: The Sensitivity Function in Variability Analysis. IEEE Trans. Reliability R-15, Aug. 1966, pp. 70-76.

8. Gray, K. B., Jr.: Worst Distribution Analysis for Statistical Circuit Design. ASQC 21 Tech. Conf. Proceedings, 1967, pp. 207-211.

9. Kendall, M. G.: The Advanced Theory of Statistics. Vol. II. Griffin & Co. Ltd., 1948.

10. Draper, N.; and Smith, H.: Applied Regression Analysis. John Wiley & Sons, Inc., 1966.

11. Pearson, E. S.; and Hartley, H. O.: Biometrika Tables for Statisticians. Vol. I. Cambridge Uni. Press, 1956

12a. Sarhan, A. E.; and Greenberg, B.: Contributions to Order Statistics. John Wiley & Sons, Inc., 1962.

12b. Harter, H. L.: Estimation of parameters of life distributions. OAR Research Review, Apr. 1967, pp. 9-11.

13. Miller, I.; and Freund, J. E.: Probability and Statistics for Engineers. Prentice Hall, 1965.

14. Military Handbook 5: Metallic Materials and Elements for Flight Vehicle Structures. Rev. Feb. 1966.

15. Estes, R. H.; and Lancaster, E. R.: Two-point Taylor Series Expansions. NASA TMX-55759 (X-643-66-657, N67-23965) Goddard Space Flight Center, Greenbelt, Md., December 1966.

16. Shooman, M. L.: Mathematical Models for Drift Failure Analysis. Proc. 11th Nat'l Symposium Reliability and Quality Control. Jan. 1965, pp. 557-567. (available from IEEE hdqtrs.)

17. Rao, C. R.: Advanced Statistical Methods in Biometric Research. John Wiley & Sons, Inc., 1952.

Appendix A

Derivation of Mean and Variance Errors Due to a Pip on the Tail of a Normal Distribution

In estimating a true mean or variance from a sample, no matter how large, there is always an uncertainty in the estimate due to the finite sample size. This appendix uses that uncertainty and compares it to the change introduced by a small pip way out on the tail of the distribution. The distribution being analyzed consists of a Normal* distribution (mean μ, variance σ^2) with area (weight) of unity and a pip of area (weight) $f << 1$ at $\pm k\sigma$ on the right tail. The prime refers to the combined distribution.

The mean is found by first moments:

$$\mu' \times (1 + f) = \mu \times 1 + (\mu + k\sigma) \times f$$

$$\text{or } \Delta\mu \equiv \mu' - \mu = k\sigma f/(1 + f) \simeq k\sigma f$$

The variance is found by second moments:

$$\sigma'^2(1 + f) = (\sigma^2 + \overline{\Delta\mu}^2) \times 1 + (k\sigma - \Delta\mu)^2 \times f$$

$$\text{or } \Delta\sigma^2 \equiv \sigma'^2 - \sigma^2 = \left(\frac{k}{1+f}\right)^2 f(1 + f)\,\sigma^2 \simeq k^2 f\sigma^2$$

If $\hat{\mu}$, $\hat{\mu}'$, $\hat{\sigma}^2$, $\hat{\sigma}'^2$ are estimates** from a single sample of N, they will have distributions whose standard deviations (sd) are:

$$sd(\hat{\mu}) \simeq sd(\hat{\mu}') \simeq \sigma/\sqrt{N}$$

$$sd(\hat{\sigma}'^2) \simeq \sigma^2/\sqrt{\tfrac{1}{2}N}$$

Now, use the standard deviation as the measure of uncertainty in an estimate. Then set the uncertainty in the estimate equal to the change due to the pip. At this point the change can "just" be detected. A smaller sample with its larger uncertainty would not detect the changes due to the pip; a larger sample with its smaller uncertainty could detect those changes due to the pip.

$$sd(\hat{\mu},\hat{\mu}') \simeq \Delta\mu, \text{ i.e., } \quad k\sigma f = \sigma/\sqrt{N_1} \;, \quad N_1 = 1/k^2f^2 \;.$$

$$sd(\hat{\sigma}^2,\hat{\sigma}'^2) \simeq \Delta\sigma^2, \text{ i.e., } \quad k^2 f\sigma^2 = \sigma^2\sqrt{2/N_2} \;, \quad N_2 = 2/k^4f^2 = 2N_1/k^2 \;.$$

For $k > \sqrt{2}$, $N_2 < N_1$ and the change in variance can be detected before the change in mean.

---

* Normality is sufficient in this example, but not always necessary. For the sample mean, the assumption is not necessary. For the variance, it may not be necessary since the sample size is so large. In both cases, the sample size is large enough so that the standard deviation is a good measure of the uncertainty.

** The circumflex or "hat" over a parameter name refers to an estimate from the data, of that parameter.

Appendix B

Taylor's Series*

Any reasonably well behaved function can be expanded in a Taylor's series; one of the obvious criteria being that the function and all needed derivatives be defined at the point of expansion. Virtually all functions used in engineering analysis which meet this obvious criterion will be well behaved enough. There are many equivalent forms of Taylor's series used in the literature; they revolve around how the variables are expressed (e.g., $x+h$, $x-x_o$), if there is a remainder, and if so, how the remainder is written. The remainder form is used here because it allows estimation of the error when the series is truncated, as it must be in numerical analysis. The Lagrangian form of the remainder is adopted because it is easy to write, use, and understand.

For one variable:

$$f(x) = f(x_o + h) = f(x_o) + hf'(x_o) + \tfrac{1}{2}h^2 f''(x_o) + \ldots + \frac{1}{N!} h^N f^{(N)}(x_o + \theta h) \ .$$

$$= \sum_{i=o}^{N} \frac{1}{i!} h^i \left(\frac{d^i f(x)}{dx^i}\right)\Bigg|_{x = \begin{cases} x_o & , \ i \neq N \\ x_o + \theta h, & i = N \end{cases}}$$

For two variables:

$$f(x,y) = f(x_o + h, \ y_o + k) = f(x_o,y_o) + h \ f_x(x_o,y_o) + k \ f_y(x_o,y_o)$$

$$+ \tfrac{1}{2}h^2 \ f_{xx}(x_o,y_o) + h \ k \ f_{xy}(x_o,y_o) + \tfrac{1}{2}k^2 \ f_{yy}(x_o,y_o) + \ldots$$

$$+ \sum_{j=o}^{N} \frac{h^j \ k^{N-j}}{j!(N-j)!} \ \frac{\partial^N f(x,y)}{\partial x^j \ \partial y^{N-j}}\Bigg|_{\substack{x = x_o + \theta h \\ y = y_o + \theta k}}$$

$$= \sum_{i=o}^{N} \frac{1}{i!} \sum_{j=o}^{i} \binom{i}{j} h^j \ k^{i-j} \left(\frac{\partial^i f(x,y)}{\partial x^j \ \partial y^{i-j}}\right)\Bigg|_{x,y = \begin{cases} x_o,y_o; \ i \neq N \\ x_o + \theta h, \ y_o + \theta k; \ i = N \end{cases}}$$

---

* The notation is described at the end of the appendix.

A general formula for n variables is:

$$f(\underline{x}) = f(\underline{x}_o + \underline{h}) = \sum_{o}^{N} \frac{1}{i!} \left(\sum_{i}^{n} h_j \frac{\partial}{\partial x_h} \right)^i f(\underline{x}) \Bigg|_{\underline{x} = \begin{cases} \underline{x}_o, i \neq N \\ \underline{x}_o + \theta\underline{h}, \; i = N \end{cases}}$$

$$= \left(\exp \sum_{j=1}^{n} h_j \frac{\partial}{\partial x_j} \right) f(\underline{x}) \Bigg|_{\underline{x} = \underline{x}_o}$$

The first equation uses the Lagrange form of the remainder (the N-th term is the remainder) in which it is asserted that there exists at least one θ such that the equality is satisfied. The second equation is a more compact way of writing the formula and is perhaps easier to remember; it does not use the remainder form. The nomenclature is quite symbolic; it is interpreted by using the power series expansion for the exponential and applying it term by term. A similar remark is true where partial derivative operators $(\frac{\partial}{\partial x})$ are raised to a power. Just perform the indicated algebra, treating the operator as you would any other symbol.

It is important to realize that all the derivatives are evaluated at a particular point as shown. They do not appear as continuous functions.

A Taylor's series has the advantage of a power series in that it can be differentiated term-by-term to get the series for the derivative of a function. It has a further advantage that the coefficients do not depend on the length of the series.

It is not to be presumed that a Taylor's series expansion is always the best expansion for an N-1 term expansion of a function. Other methods include expansion in terms of orthogonal functions such as Fourier series (trigonometric functions), Hermite polynomials, and Chebyshev polynomials. Each has its advantages. One may be a "least-squares" solution, another may have its actual error bounded, etc. A numerical analyst, or equivalent, can be helpful in deciding which expansion to use if more than the first term is needed. Most of them are equivalent up to the first term.

Formulas are in existence for the expansion of f(x) about 2 points, but they are rarely used. See Ref. 15.

Notation:

$$f^{(N)}(x_o) \equiv \frac{d^N f(x)}{dx^N} \Bigg|_{x=x_o}$$

expression $\Big|_{x=a}$ means evaluate the expression at x = a.

$$f_{xy}(x_0,y_0) \equiv \frac{\partial^2 f(x,y)}{\partial x \partial y}\Bigg|_{x,y=x_o,y_o} .$$

$$\binom{n}{m} \equiv \frac{n!}{m!(m-n)!} = \text{binomial coefficient.}$$

$\underline{x}$ implies $x_1, x_2, \ldots, x_n$; $\underline{h}$ implies $h_1, h_2, \ldots, h_n$;

$\underline{x}_0 + \underline{h}$ implies $x_{01} + h_1, x_{02} + h_2, \ldots, x_{0n} + h_n$.

There exists at least one θ $(0 < \theta < 1)$ such that the equations are true.

$\left(\frac{\partial}{\partial x}\right)^n f(x)$ implies $\frac{\partial^n f(x)}{\partial x^n}$

Appendix C

Mean Value Theorem

The mean value theorem can be considered a special case of a Taylor's series expansion with the Lagrangian form of the remainder for the series. From Appendix B we have*

$$f(\underline{x}_o + \underline{h}) = \sum_{i=o}^{N} \frac{1}{i!} \left(\sum_{j=1}^{N} h_j \frac{\partial}{\partial x_j} \right)^i f(\underline{x}) \Bigg|_{\underline{x} = \begin{cases} \underline{x}_o, & i \neq N \\ \underline{x}_o + \theta\underline{h}, & i = N \end{cases}}$$

If N=1, this reduces to the mean value theorem for n dimensions (variables)

$$f(\underline{x}_o + \underline{h}) = f(\underline{x}_o) + h_1 f_1 (\underline{x}_o + \theta\underline{h}) + h_2 f_2 (\underline{x}_o + \theta\underline{h}) +$$

$$\ldots + h_n f_n (\underline{x}_o + \theta\underline{h}) ,$$

where $f_i \equiv \partial f / \partial x_i$.

For one dimension (variable) this reduces to the familiar form

$$f(x_o + h) = f(x_o) + h f'(x_o + \theta h) .$$

---

* There exists at least one θ $(0 < \theta < 1)$ such that the equation is true. The notation $\underline{x}$ implies $x_1, x_2, \ldots x_m$ as before. The notation $\underline{x}_o + \underline{h}$ implies $x_{01} + h_1$, $x_{02} + h_2$, ..., $x_{0m} + h_m$.

APPENDIX D

Partial Differentiation

Partial differentiation is no more difficult than ordinary differentiation--although the former is often more tedious--indeed, they are both essentially the application of the same set of rules. In partial differentiation, however, some parameters are sometimes held constant and others not--the trick is to know "which ones" and "when".

Notation is important here if the situation is at all complicated. The designation "partial" by itself doesn't mean anything except as the set of variables-to-be-treated-as-independent is given or easily inferred. In complicated situations it is better not to leave anything for easy inference. Two notations are commonly used:

(1) List all the variables-to-be-treated-as-independent with each differentiation.

(2) List only the variables being held constant for this differentiation.

Obviously, the second, plus the variable used for differentiation, is the first. The notation used in this volume on the basis of simplicity and of ease of writing is the first, viz., list all the variables-to-be-treated-as-independent.

Let $\underline{f} \equiv \underline{f}(\underline{x})$ and $\underline{x} \equiv \underline{x}(\underline{s})$,* the formula for a partial derivative is then

$$\left.\frac{\partial f_k}{\partial s_j}\right|_{\underline{s}} = \sum_i \left.\frac{\partial f_k}{\partial x_i}\right|_{\underline{x}} \times \left.\frac{\partial x_i}{\partial s_j}\right|_{\underline{s}}$$

The potential complications become more deep when some of the x_i are identical to some of the s_j and perhaps some of the f_k are identical to some s_j and/or x_i. This can happen when several equations are written to express the relationships between many parameters. The way to proceed is to be very careful of the notation and to realize that if $y = s_1 = x_1$, then $\partial f/\partial y|_{\underline{x}}$ and $\partial f/\partial y|_{\underline{s}}$ are different both in physical meaning and in mathematical formulation.

---

* $\underline{f}(\underline{x})$ implies $f_1(\underline{x}), f_2(\underline{x}), \ldots, f_m(\underline{x})$
$\underline{x}$ implies $x_1, x_2, \ldots, x_n$
$\underline{s}$ implies $s_1, s_2, \ldots, s_p$

APPENDIX E

Central Limit Theorem

There are many variations of statements of the Central Limit theorem, depending on how complicated a proof it is desired to use. The stronger statements are less restrictive on the nature of the variables and usually require more proof. A wording which is usually sufficient for reliability purposes is the following: If

(1) x_i $(i = 1, \ldots, n)$ are statistically independent variables, each with variance σ_i^2;

(2) $y_n \equiv \sum_1^n x_i$, $S_n^2 \equiv \sum_1^n \sigma_i^2$, $\bar{y}_n \equiv$ average (expected) value of y_n;

(3) $Z_n \equiv (y_n - \bar{y}_n)/S_n$

(4) No fixed subset of σ_i^2 forms a substantial part of S_n^2 as n becomes large.

Then, in the limit, as $n \to \infty$: Z_∞, if it exists, has a Normal distribution with zero mean and unit variance.

Unfortunately nothing is said in this theorem about how large n must be to have the distribution close enough to the Normal. It is reasonable to presume that n must be greater when

(1) one goes further out in the tails of the distribution,

(2) the less Gaussian-like the variables are,

(3) the more accuracy is desired.

It is not difficult to dream up distributions for individual x_i which are pathological enough to cause no end of trouble in the Central Limit theorem, i.e., they would require n to be very large before any semblance of Normality is obtained.

Before invoking the Central Limit theorem make sure that it is necessary to an important argument to do so, and then be very careful about jumping from the theorem itself to the result for a finite number of variables and way out on the tails of the distribution.

Appendix F

Tractable Combinations of pdf's[a] *

This list is not complete since what is or is not tractable is a matter of opinion. Also some rather simple distributions with no convenient formulas are not included. In all but (1) and (10b) a sufficient condition on the x's is that they be statistically independent (in (1) and (10b) the variables can be statistically dependent). While it may not always be a necessary condition, the analysis is too complicated in most cases for anyone to tell.

The distributions mentioned, along with some of their properties, are given in Appendix G for convenient reference. The parameters referred to in columns 2 and 3 are those given in Appendix G--other references may use different notation.

Linear combination of variables[b]	Original distribution	Final distribution[b]
(1) $y = \sum a_i x_i$	x_i is Normal (Gaussian)	y is Normal (Gaussian) (see Sec. 5.2.1 for new parameters)
Sum of statistically independent variables[b],[c]		
(2) $m = \sum n_i$	n_i is Poisson with parameter μ_i	m is Poisson with parameter $\sum \mu_i$
(3a) $y = \sum \chi_i^2$	χ_i^2 is Chi-square with parameter ν_i	y is Chi-square with parameter $\sum \nu_i$
(3b) $y = \sum x_i^2$	x_i is Standard Normal	y is Chi-square with parameter $\nu = n$
(4a) $y = \sum x_i$	x_i is Gamma (incomplete) with parameters α_i, β	y is Gamma (incomplete) with parameters $\sum \alpha_i$, β
(4b) $y = \sum x_i$	x_i is exponential with parameter λ; i.e., from (4a), $\alpha_i = 1$, $\beta = \lambda$	y is Gamma (incomplete) with parameters n, λ
(5) $y = \sum x_i$	x_i is Cauchy with parameters α, β	y is Cauchy with parameters $\alpha, n\beta$

* The notes cited by (a), (b), (c), etc. are listed at the end of the section.

	Original distribution	Final distribution
Least of statistically independent variables[c]		
(6) y = least of x_i	x_i is Weibull with parameters α_i, β	y is Weibull with parameters $[\sum (\alpha_i)^{-\beta}]^{-1/\beta}$, β
(7) y = least of x_i	x_i is exponential with parameter λ_i; i.e., from (6), $\alpha_i = 1/\lambda_i$, $\beta = 1$	y is exponential with parameter $\sum \lambda_i$
Product of statistically independent variables[b][c]		
(8) $y = \prod x_i$	x_i is Beta (incomplete) with parameters a_i, b_i such that $a_i = a_{i+1} + b_{i+1}$	y is Beta (incomplete) with parameters a_n, $\sum b_i$
Quotient involving two statistically independent variables[c][e]		
(9) $y = \dfrac{1}{1 + \dfrac{x_1}{x_2}}$	x_i is Gamma (incomplete) with parameters α_1, β, and α_2, β	y is Beta (incomplete) with parameters α_1, α_2
(10a) $y = x_1/x_2$	x_i is Standard Normal	y = is Cauchy with parameters $\alpha = 0$, $\beta = 1$
(10b)[e] $y = \dfrac{x_1}{x_2} \cdot \dfrac{1}{\varepsilon}$	x_i is Standard Normal[f]	y is Cauchy with parameters $\alpha = \rho/\varepsilon$, $\beta = 1$
(11)[d] $y = \dfrac{x}{\sqrt{\chi^2/\nu}}$	x is Standard Normal; χ^2 is Chi-square with parameter ν	y is Student's t with parameter ν

(a) Some of these were adapted from Refs. 16 and 17.

(b) $\sum$ implies $\sum_1^n{}_i$, $\prod$ implies $\prod_{1i}^{n}$

(c) See text above

(d) χ^2 and x need not come from the same sample. This formula is quite general and nothing more is meant for restrictions than is shown.

(e) In (10b), the variables need not be statistically independent.

(f) The linear correlation coefficient is ρ, ($\varepsilon^4 = 1 - \rho^2$).

Appendix G

Some Probability Distributions[a]

Name	pdf[b][c]	cdf(from right)[b][c]	mean[d]	variance[d]	coef. of skewness[d]	domain of variables (x,p,χ^2) and parameters[b][e] $(\alpha,\beta,\lambda,\nu)$
Normal (Gaussian)	$\frac{1}{\sqrt{2\pi}\,\sigma}\exp\{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\}$,	$\bar{\Phi}(\frac{x-\mu}{\sigma})$ or $G(\frac{x-\mu}{\sigma})$	μ	σ^2	0	$-\infty < x < \infty$ $-\infty < \mu < \infty$ $0 < \sigma < \infty$
Standard Normal	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$	$\Phi(x)$ or $G(x)$	0	1	0	$-\infty < x < \infty$
exponential[f]	$\lambda\exp(-\lambda x)$	$\exp(-\lambda x)$	$1/\lambda$	$1/\lambda^2$	2	$0 \le x < \infty$ $0 < \lambda < \infty$
Weibull[g]	$\frac{\beta}{\alpha}(\frac{x}{\alpha})^{\beta-1}\exp\{-(\frac{x}{\alpha})^{\beta}\}$,	$\exp\{-(\frac{x}{\alpha})^{\beta}\}$	$\alpha\Gamma(\frac{1}{\beta}+1)$	$\alpha^2[\Gamma(\frac{2}{\beta}+1)-\Gamma^2(\frac{2}{\beta}+1)]$, (h)		$0 \le x < \infty$ $0 < \alpha < \infty$ $0 < \beta < \infty$
log Normal	change the variable y to x = ln y and use the Normal distribution above. This is not the same as taking y directly as the variable, but in reliability work is usually as good or better. (Log_{10} y can be used instead of ln y if the variance is changed.)					
(incomplete) Gamma[p]	$\frac{\beta}{\Gamma(\alpha)}(\beta x)^{\alpha-1}e^{-\beta x}$	(i) $\frac{\Gamma(\alpha,\beta x)}{\Gamma(\alpha)}$	α/β	α/β^2	$\frac{2}{\sqrt{\alpha}}$	$0 \le x < \infty$ $0 < \alpha < \infty$ $0 < \beta < \infty$
(incomplete) Beta[j]	$\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$,	$1 - I_x(\alpha,\beta) = I_{1-x}(\beta,\alpha)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\frac{2(\alpha-\beta)}{\alpha+\beta+2}$	$0 \le x \le 1$ $0 < \alpha < \infty$ $0 < \beta < \infty$
Chi-square[k]	$\frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2}\nu)}(\frac{1}{2}\chi^2)^{(\frac{1}{2}\nu-1)}e^{-\frac{1}{2}\chi^2}$,	$Q(\chi^2\|\nu) = \frac{\Gamma(\frac{1}{2}\nu,\frac{1}{2}\chi^2)}{\Gamma(\frac{1}{2}\nu)}$	ν	2ν	$\frac{1}{\sqrt{\nu/8}}$	$0 \le \chi^2 < \infty$ ν = positive integer

Name	pdf[b][c]	cdf (from right)[b][c]	mean[d]	variance[d]	coef. of skewness[d]	domain of variables (x,p,χ^2) and parameters[b][e] $(\alpha,\beta,\lambda,\nu)$
Cauchy[1]	$\frac{1}{\pi\beta}\,\frac{1}{1+(\frac{x-\alpha}{\beta})^2}$	$\frac{1}{\pi}\cot^{-1}(\frac{x-\alpha}{\beta})$	(m)	(m)	(m)	$-\infty < x < \infty$ $-\infty < \alpha < \infty$ $0 < \beta < \infty$
Student's t	$\frac{\Gamma(\frac{1}{2}\nu+\frac{1}{2})}{\sqrt{\pi\nu}\,\Gamma(\frac{1}{2}\nu)}(1+\frac{t^2}{\nu})^{-(\frac{1}{2}\nu+\frac{1}{2})}$	$Q(t\|\nu)$	0	$\frac{\nu}{\nu-2}$ (n)	0	$-\infty < x < \infty$ ν = positive integer
	point probability[c] (n-th term)	cumulative sum[c] (from right)				domain of variables (n = integer) and parameters[b] (μ,p,N)
Poisson	$e^{-\mu}\,\frac{\mu^n}{n!}$	$1-Q(\chi^2=2\mu\|\nu=2n>0)$	μ	μ	$\frac{1}{\sqrt{\mu}}$	$0 \le n < \infty$ $0 < \mu < \infty$
Binomial[o]	$\binom{N}{n}p^n q^{N-n}$	$I_p(n, N-n+1)$	Np	Npq	$\frac{q-p}{\sqrt{Npq}}$	$0 \le n \le N$ $0 < p = 1-q < 1$ N = positive integer

(a) From Refs. 3 and 17.

(b) All parameters and variables are presumed real, although this is not always a necessary restriction mathematically. The basic restriction on the parameters is that both the cdf and pdf exist. The restrictions given here are sufficient, but may not be necessary.

(c) pdf is probability density function; cdf is cumulative distribution function. Two good sources for definitions, explanations, uses, and tabulations are Refs. 3 and 11. The hazard rate is pdf/cdf.

(d) This does not imply that the mean, variance, and skewness coefficients are the best measures of central tendency, dispersion, and nonsymmetry respectively nor that they are the only functions of interest. Mean is the first moment about the origin. Variance (square of the standard deviation) is the second moment about the mean. Coefficient of skewness is the ratio: third-moment-about-the-mean/cube-of-standard-deviation, and is dimensionless. They are useful chiefly for unimodal distributions and, where skewness is not given, only for reasonably symmetric ones.

(e) When $x \geq 0$, a "guarantee period" x_o can be introduced. This results in (a) substituting $x-x_o$ for x, (b) adding x_o to the mean, (c) changing "$0 \leq x < \infty$" to "$x_o \leq x < \infty$", (d) adding $-\infty < x_o < \infty$, and (e) noting that there is an extra parameter in the function.

(f) This is the Weibull distribution with $\alpha = 1/\lambda$, $\beta = 1$, or the (incomplete) Gamma distribution with $\alpha = 1$, $\beta = \lambda$.

(g) There are other equivalent forms. This has the advantage that x and α have the same physical units.

(h) $\{\Gamma(\frac{3}{\beta}+1) - 3\Gamma(\frac{2}{\beta}+1)\,\Gamma(\frac{1}{\beta}+1) + 2\Gamma^3(\frac{1}{\beta}+1)\}/\{\Gamma(\frac{2}{\beta}+1) - \Gamma^2(\frac{1}{\beta}+1)\}^{3/2}$.

(i) Tables for $\Gamma(\alpha,\beta x)$ are rare. But for $\alpha = \frac{1}{2}\nu$ and $\beta x = \frac{1}{2}\chi^2$, $\frac{\Gamma(\frac{1}{2}\nu,\ \frac{1}{2}\chi^2)}{\Gamma(\frac{1}{2}\nu)}$ is $Q(\chi^2|\nu)$

(j) $B(\alpha,\beta) \equiv \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function.

(k) This is the (incomplete) Gamma with $\beta = \frac{1}{2}$, $\alpha = \frac{\nu}{2}$, $x = \chi^2$.

(l) This is the Student's t distribution with $\nu = 1$, $t = \frac{x-\alpha}{\beta}$.

(m) These involve integrals whose values are "infinite" and thus not defined. It is tempting to point out that α is the center of symmetry.

(n) For $\nu = 1,2$, these involve integrals whose values are "infinite" and thus not defined.

(o) $\binom{N}{n} \equiv \frac{N!}{n!(N-n)!}$ is the binomial coefficient.

(p) $\Gamma(\alpha)$ is the gamma function. $\Gamma(n) = (n-1)!$ for n a positive integer. $\Gamma(\alpha)$ is reasonably well tabulated for α not an integer. The Incomplete Gamma distribution is a special case of the Pearson type III distribution.

NASA-Langley, 1968 — 15 CR-1126